MEASUREMENT IN HEALTH PSYCHOLOGY

EDITED BY: Paola Gremigni, Antonio De Padua Serafim, Giulia Casu and Victor Zaia

PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-378-8
DOI 10.3389/978-2-88976-378-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

MEASUREMENT IN HEALTH PSYCHOLOGY

Topic Editors:
Paola Gremigni, University of Bologna, Italy
Antonio De Padua Serafim, University of São Paulo, Brazil
Giulia Casu, University of Bologna, Italy
Victor Zaia, Faculdade de Medicina do ABC, Brazil

Citation: Gremigni, P., De Padua Serafim, A., Casu, G., Zaia, V., eds. (2022). Measurement in Health Psychology. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88976-378-8

Table of Contents

frontiers | Frontiers in Psychology

EDITORIAL
published: 24 May 2022
doi: 10.3389/fpsyg.2022.937700

Editorial: Measurement in Health Psychology

Giulia Casu[1], Antonio de Padua Serafim[2], Victor Zaia[3]* and Paola Gremigni[1]

[1] Department of Psychology, University of Bologna, Bologna, Italy, [2] Institute of Psychology, University of São Paulo, São Paulo, Brazil, [3] Postgraduate Program in Health Sciences, Centro Universitário FMABC, Santo André, Brazil

Keywords: Health Psychology, measurement, health outcomes, validation, psychometrics

OPEN ACCESS

Edited and reviewed by:
Pietro Cipresso,
University of Turin, Italy

Correspondence:
Victor Zaia
victorzaia@gmail.com

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *06 May 2022*
Accepted: *10 May 2022*
Published: *24 May 2022*

Citation:
Casu G, de Padua Serafim A, Zaia V and Gremigni P (2022) Editorial: Measurement in Health Psychology. Front. Psychol. 13:937700. doi: 10.3389/fpsyg.2022.937700

Editorial on the Research Topic

Measurement in Health Psychology

According to the American Psychological Association (APA, 2003; Freedland, 2021), Health Psychology is an interdisciplinary subspecialty of Psychology concerned with the study of biological, behavioral, and social factors contributing to both health and illness. Health Psychology applies its principles, techniques, and scientific knowledge to evaluate, diagnose, treat, modify and prevent physical, mental, or any other problems relevant to the processes of health and disease. It focuses on the promotion and maintenance of health, the prevention and treatment of illness, and the identification of etiologic and diagnostic correlates of health, disease, and dysfunctions.

Several issues and factors are still understudied in Health Psychology. Psychological effects of the COVID-19 outbreak, caregiving experiences, work-related stress and life-work balance, quality of life of sexual minorities, health-related orientations and motivations among younger generations, and the impact of the new media on people's mental health are just a few examples of Health Psychology phenomena that need further investigation. A critical issue in Health Psychology is measurement. A variety of crucial psychological constructs, health-related behaviors, and responses to health, illness, and healthcare need to be addressed and measured in Health Psychology research and practice. Intervention programs designed to foster good health by changing negative health behavior, promoting positive health behavior, and enhancing the management of chronic conditions have been of particular interest to Health Psychology since its beginnings as a distinctive discipline. The availability of psychometrically sound measures is crucial to assessing the effectiveness of such interventions and deepening our understanding of the social and psychological processes of health and illness (Apple, 2005; Etches et al., 2006).

We launched the Research Topic "Measurement in Health Psychology" in this context. Because measures are a critical part of research and practice in Health Psychology, we were interested in providing an overview of up-to-date measurement principles and methods, which could help improve the process of developing valid, reliable, and sensitive instruments to be used in the fields of Health Psychology.

Within this Research Topic, we brought together research studies developing new tools or advancing the psychometric study of existing measures relevant to Health Psychology. Twenty-three articles authored by 105 contributors from different countries and continents were accepted. The research contributions published in this Research Topic relate to self-report measures of socio-environmental and psychological/behavioral influences on health and disease, as framed within the biopsychosocial model (Engel, 2012), which has traditionally guided the field of Health Psychology, and the study of health outcomes.

SOCIO-ENVIRONMENTAL INFLUENCES ON HEALTH AND DISEASE

Six articles in this Research Topic addressed constructs belonging to the social domain of the biopsychosocial model. Three papers dealt with the assessment of family relationships. One of them (Zhan and Wang) presented the development of a new measure to assess the subjective evaluation of harmonious family relations as a resource for physical and mental health. Three independent samples of Chinese university students were enrolled, respectively, for item analysis and exploratory factor analysis (EFA) of the initial item pool; confirmatory factor analysis (CFA) and assessment of internal consistency; and relationships to other variables and test-retest reliability over a two-month interval. Results supported a nine-factor model of family harmony, with adequate reliability and validity as evidenced by positive correlations with criterion measures of subjective well-being and family function, cohesion, and adaptability, and negative correlations with loneliness scores.

Another paper (Guo et al.) evaluated the psychometric properties of a Chinese version of the 10-item Family Communication Scale (FCS) to measure positive family communication. EFA and CFA supported a one-factor structure, which showed adequate internal consistency and test-retest reliability and correlated in the expected direction with wellbeing indicators and frequency of communications with family members via Information and Communication Technologies.

Finally, UK researchers (Bywater et al.) proposed a new short measure to assess the bonding between parents and children under 1 year of age, namely the "Me and My Baby" (MaMB) questionnaire. Factor analyses and Rasch calibration performed on data from 434 mothers provided initial evidence that the MaMB reliably measures infant bonding.

In the context of caregiving, one paper tested the psychometric characteristics of the Zarit Burden Interview in Peruvian informal primary caregivers of persons diagnosed with intellectual disabilities (Boluarte-Carbajal et al.). Applying CFA and Rasch analysis, the authors found evidence of a unidimensional structure with adequate reliability. Evidence of validity was provided by relationships with a measure of the risk of physical and psychosocial abuse and neglect by primary caregivers. As for formal caregiving, Italian researchers developed a 20-item scale to assess emotional, informational, appraisal, and instrumental social support by healthcare providers in the oncology setting (Tomai and Lauriola). Scale dimensionality and reliability were tested using exploratory structural equation modeling (ESEM) and Mokken scaling analysis. Evidence of validity and reliability and expected associations with doctor communication skills and trust in the physician supported the use of this new measure of healthcare social support as multidimensional a construct.

Finally, as for the school environment, one article (Carmona-Halty et al.) tested the validity and reliability of the School Burnout Inventory among Chilean high school students. CFAs on the 8-item Chilean version supported two statistically equivalent first- and second-order three-factor models of school burnout as composed of exhaustion, cynicism, and inadequacy, which showed gender invariance. Internal consistency and test-retest reliability were acceptable, and school burnout scores correlated as expected with study-related emotions, academic psychological capital, and academic engagement.

PSYCHOLOGICAL/BEHAVIORAL INFLUENCES ON HEALTH AND DISEASE

Nine papers addressed the assessment of individual characteristics and behaviors relevant to health and illness. Within the field of psychosomatic disorders, Iranian researchers (Lashkari et al.) tested the psychometric properties of the Farsi version of Perth Alexithymia Questionnaire (PAQ). Based on CFA results, the 5-factor model of the original PAQ replicated well on data from college students, with adequate internal consistency and test-retest reliability. Evidence of associations with other variables (TAS-20, emotion regulation, depression, anxiety) was collected for the Farsi PAQ.

The Work-Related Rumination Scale was tested on a sample of Puerto Rican workers (Rosario-Hernandez et al.). An 11-item Spanish version with three factors (affective rumination, problem-solving pondering, and detachment) was obtained using CFA and ESEM and proved to be invariant across gender and age within and between five different study samples. Reliability coefficients were satisfactory, and correlations with relevant criterion measures (e.g., sleep quality, emotional exhaustion) provided evidence of convergent and divergent validity.

As to individual beliefs and expectations, one article (Lang and Ye) presented a Chinese adaptation of the Self-Objectification Beliefs and Behaviors Scale (C-SOBBS). As a result of exposure to sexual objectification in interpersonal situations and visual media, self-objectification entails viewing one's own body from a third person's perspective and has been linked to poorer women's mental health. The authors found evidence of structural (CFA), convergent, discriminant, and incremental validity, and adequate internal consistency and test-retest reliability for the C-SOBBS. In another paper, Chinese researchers proposed the development and validation of the Psychological Needs of Cancer Patients Scale to identify the psychological care demands of cancer patients (Chen et al.). Results of EFA and CFA supported a six-factor model of value and esteem, independence and control, mental care, disease care, belonging and companionship, and security, with acceptable reliability and expected associations with anxiety and depression.

Two articles addressed situation-specific coping. The Robust Pandemic Coping Scale (R-PCS) was developed to assess coping strategies related to pandemic situations at all stages of the epidemic management cycle (Burro et al.). Data from Italian university students were analyzed via EFA and CFA, followed by Rasch analysis. A four-factor model of despair, adjustment, proactivity, and aversion was supported, which was invariant across gender and age and showed adequate reliability. Discriminant and criterion-related validity based on correlations with personality characteristics helpful in coping with disasters and predictive validity on levels of enjoyment and anger 2 months later were also supported. To assess coping strategies to deal with cancer, researchers from Portugal (Lemos et al.) performed a cross-cultural adaptation and psychometric evaluation of the Perceived Ability to Cope with Trauma Scale (PACT). Results of CFA on data from patients recently diagnosed with early breast cancer supported the original PACT two-factor model of coping flexibility as composed of forward and trauma focus domains, which showed adequate internal consistency and associations in the expected direction with self-efficacy to cope with cancer, quality of life and psychological distress.

In the area of health-related attitudes, Chilean researchers (Ferrer-Urbina et al.) developed a brief scale to assess the affective, cognitive, and behavioral attitudes of youth and young adults toward condom use. Using EFA and ESEM, the authors found support for the hypothesized three-factor model. The scale showed strong invariance across gender, adequate reliability, and expected relationships with sexual risk behaviors and condom use.

Among health-related behaviors, one study (Milasauskiene et al.) tested the psychometric performance of a very brief measure of problematic internet use, namely the Compulsive Internet Use Scale (CIUS), when used with Lithuanian medical students and resident doctors. Results indicated that the brief, 5-, 7-, and 9-item versions of the CIUS were reliable and valid screening tools to assess the severity of symptoms of problematic internet use in the medical population. Another brief, 6-item scale of excessive use of social networks was adapted for use with Mexican adolescents and young adults (Salas-Blas et al.). Altogether, the structural properties of the response options fitted the partial-credit model. CFA supported a single domain of addiction to social networks, with measurement invariance across sex, age, and educational campus, good internal consistency, and theoretically consistent associations with sensation seeking and depression.

HEALTH OUTCOMES

A total of eight papers focused on assessing outcomes of the interactive relationships between biological, environmental, and psychological/behavioral factors. Two papers focused on health-related quality of life (HRQoL), which is a primary outcome in the evaluation of interventions' effectiveness on people's health. One article (Xu et al.) presented the traditional Chinese version of the Recovering Quality of Life (ReQoL) outcome measure. The ReQoL measures mental health recovery, defined as a self-directed process of healing and transformation, which has received increasing attention in evaluating the outcomes of mental care. The Chinese ReQoL showed good psychometric properties in terms of internal consistency, test-retest reliability, factor structure (CFA), known-group validity, and associations with relevant variables in the general population.

Lithuanian researchers (Gecaite-Stonciene et al.) examined the validity and reliability of the Minnesota Living with Heart Failure Questionnaire in individuals with coronary artery disease. EFA and CFAs supported a three-factor model of physical, social, and emotional disease-specific HRQoL. The Lithuanian version was reliable and showed evidence of convergent validity based on correlations with another measure of HRQoL (SF-36) and with exercise capacity assessed using a standardized computer-driven bicycle ergometer.

Six papers addressed the assessment of indicators of mental health or adjustment to health problems. One of them presented the validation of a Kazakhstani version of the Mental Health Continuum–Short Form to assess emotional, social, and psychological wellbeing (Hernandez-Torrano et al.). CFA confirmed a bifactor model, which was invariant across gender and age in a sample of university students. Based on reliability analyses, the authors advised against the interpretation of specific-factor scores and recommended the computation of a general wellbeing score.

Another article (Flenreiss-Frankl et al.) presented the validation study of a multidimensional inventory for the assessment of mental pain after traumatic experiences (FESSTE30) in the German speaking general population. CFA showed a satisfactory fit of a five-factors measuring somatization, depression, intrusive memories, dissociation, and anxiety. The scale showed evidence of reliability and strong correlations with measures of psychological distress, PTSD symptoms, and the extent of traumatic experiences.

The Swedish version of the Multidimensional Inventory for Religious/Spiritual Well-Being (MI-RSWB) was psychometrically tested to assess the spiritual wellbeing of a large sample of university students (Wenzl et al.). Based on the results of PCA and CFA, the authors proposed a revised model (MI-RSWB-R) with general religiosity and connectedness domains reflecting religious and spiritual wellbeing, respectively, and suggested using the remaining subscales (immanent and transcendent hope, forgiveness, sense of meaning, and connectedness) to gain insight into specific, separate facets of wellbeing. All MI-RSWB-R domains showed adequate reliability and relationships with the centrality of religiosity.

Brazilian and Italian researchers (Casu et al.) tested a Brazilian-Portuguese version of the 12-item Infertility-Related Stress Scale (IRSS-BP) to assess adjustment to the infertile condition. ESEM showed a bifactor model with one general and two specific intrapersonal and interpersonal stress factors, which was invariant across Brazilian and Italian infertile individuals. All three IRSS-BP factors showed adequate composite reliability and theoretically meaningful associations with gender, infertility duration, and depression scores.

Using Rasch's partial-credit model, correlation, and regression, one study (Hum et al.) tested the psychometric properties of the Starkstein Apathy Scale (SAS) to assess apathy in a sample of English speaking people having experienced a stroke. A revised 9-item version of the SAS targeting impairment of apathy/motivation was obtained, which was unidimensional and reasonably reliable, with no substantial item differential functioning across time, age, sex, and education.

Chinese researchers developed and tested a 40-item short form of the Inventory of Psychosocial Balance (CIPB-SF) to assess ego development based on Erikson's theory in Chinese older adults (Chen et al.). Ego integrity in older people is crucial to help them achieving successful aging and a better quality of life. Through a three-step process involving expert validity and piloting, item analysis and principal component analysis, CFA, and assessment of reliability and relations to criterion measures, the authors found acceptable validity and reliability evidence, supporting the use of the Chinese IPB.

CONCLUSIONS

In this Research Topic, we invited researchers from around the world to contribute to measuring key constructs in Health Psychology by sharing the development and improvement of measures to be used by Health Psychology researchers and clinicians.

Two-thirds of the accepted papers presented further development and validation of existing instruments for use in different cultural contexts or populations with specific characteristics. This emphasizes that the validation process is never complete but is instead an ongoing process, which involves collecting multiple types of evidence to support the appropriateness and meaningfulness of inferences and decisions made from measurement scores (Messick, 1989). Most papers examined the factor structure of measures using Classical Test Theory (CTT). The remaining papers used both CTT and Item Response Theory models. Within CTT, EFA and/or CFA were the most common analytic choice. However, some contributors preferred to combine the advantages of both EFA and CFA and use ESEM. About half of the articles in this Research Topic presented the validation of very brief measures, which are particularly suitable for minimizing patient and staff burden and ensuring time-efficient assessments in busy health settings.

Altogether, 23 measures with promising psychometric characteristics are made available to researchers and clinicians to ensure a valid and reliable assessment of psychosocial variables and behaviors that may influence health and illness, as well as to evaluate health outcomes and monitor the effectiveness of Health Psychology interventions. This Research Topic certainly does not exhaust the issue of measurement in Health Psychology. Still, we hope that other colleagues will take new initiatives to further share the development and testing of valid, reliable, and sensitive instruments to be used in Health Psychology research and practice.

AUTHOR CONTRIBUTIONS

All authors listed have a substantial and equal contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

The guest editors greatly appreciated the contributions received from the authors on this Research Topic and thank the reviewers for their valuable work.

REFERENCES

APA, A. P. A. (2003). *Psychology: Promoting Health and Well-Being through High Quality, Cost-Effective Treatment.* Available online at: http://www.health-psych.org

Apple, D. K. (2005). *Designing and Implementing Performance Measures Handbook*. Lisle, IL: Pacific Crest.

Engel, G. L. (2012). The need for a new medical model: a challenge for biomedicine. *Psychodyn. Psychiatry*. 40, 377–396. doi: 10.1521/pdps.2012.40.3.377

Etches, V., Frank, J., Di Ruggiero, E., and Manuel, D. (2006). Measuring population health: a review of indicators. *Annu. Rev. Public Health.* 27, 29–55. doi: 10.1146/annurev.publhealth.27.021405.102141

Freedland, K. E. (2021). Health psychology's 40th anniversary. *Health Psychol.* 40, 823. doi: 10.1037/hea0001139

Messick, S. (1989). "Validity", in *Educational Measurement*, Linn, R. L. (Ed.). New York: American Council on Education/Macmillan. p. 13–103.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 April 2021
doi: 10.3389/fpsyg.2021.658430

The Development of the College Students' Experience of Family Harmony Questionnaire (CSEFHQ)

Qisheng Zhan[1,2]* *and Qin Wang*[1]

[1] School of Education, Tianjin University, Tianjin, China, [2] Institute of Psychology, Tianjin University, Tianjin, China

The experience of family harmony, as an individual's subjective evaluation of harmonious family relations, has an important influence on the development of their physical and mental health. This study aimed to develop the College Students' Experience of Family Harmony Questionnaire that is fit for college students in China. On the basis of literature analysis and survey with questionnaires, five pairs of opposite assessment indexes were constructed in this paper, namely, Atmosphere of family (getting along vs. conflict), Responsibility to housework (undertaking housework vs. refusing housework), Time-sharing (sharing vs. self-isolatedness), Seeking help (help-seeking vs. avoidance), and Supporting family members (support-providing vs. indifference). Items of this questionnaire were collected from investigation, relevant scales, and discussion with experts. Here, 562 college students were selected for the pre-test and 696 for the formal test. The results showed that, except for the dimension of refusing housework, which has been deleted, other dimensions remain unchanged, and the final nine dimensions accounted for 66.03% of variance variation. Furthermore, the result of confirmatory factor analysis indicates that the model fit well with the data in construct validity [$\chi^2/df = 2.71$, Incremental Fit Index (IFI) = 0.90, Tucker–Lewis Index (TLI) = 0.89, Comparative Fit Index (CFI) = 0.90, root mean square error of approximation ($RMSEA$) = 0.05, standardized root mean square residual ($SRMR$) = 0.05]. The Cronbach's alpha (α) coefficient of this questionnaire was 0.97. The split-half reliability was 0.92, and the test–retest reliability was 0.75 for the total questionnaire. The total score of the questionnaire was significantly positively correlated with the total score of family function, family cohesion, family adaptability, and well-being ($r = 0.73, 0.71, 0.75, 0.51$, respectively, all $p < 0.01$), and it had a significant negative correlation with loneliness ($r = -0.56$, $p < 0.01$). The results showed that the final structure was reasonable, and reliability and validity conformed to the requirements of psychometrics. Therefore, the questionnaire developed in this study can be used as a valid instrument for assessing the experience of family harmony among college students in China.

Keywords: college students, experience of family harmony, reliability, validity, questionnaire

OPEN ACCESS

Edited by:
Antonio De Padua Serafim, Universidade de São Paulo, Brazil

Reviewed by:
Reina Granados, University of Granada, Spain
Pedro Tadeu, CI&DEI-ESECD-IPG, Portugal

***Correspondence:**
Qisheng Zhan
zqs@tju.edu.cn

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *25 January 2021*
Accepted: *08 March 2021*
Published: *12 April 2021*

Citation:
Zhan Q and Wang Q (2021) The Development of the College Students' Experience of Family Harmony Questionnaire (CSEFHQ). Front. Psychol. 12:658430. doi: 10.3389/fpsyg.2021.658430

INTRODUCTION

The value concept of family harmony and prosperity has always been valued by Chinese families. As an ideal state of family relations, family harmony refers to the harmonious coexistence in family life (Yap and Tan, 2011). Family harmony has a very important impact on individual mental health and happy life. Relevant studies show that college students with harmonious families have a higher level of trust than those with quarrelsome families (Zhang, 2016). College students with strong family cohesion and less conflicts show better academic, social, and emotional adaptability during college (Johnson et al., 2010). College students who experienced more family conflict reported more psychological and emotional distress, more depressive symptoms, and poorer social adjustment (Hannum and Dvorak, 2004; Lucas-Thompson and Hostinar, 2013; Rhoades and Wood, 2014). However, the evaluation of family harmony cannot be generalized. Depending on age, experience, personality, and other factors, everyone's views on family harmony may be different. Numerous studies have shown that parents and adolescents may view family relationships in different ways and have a different understanding of family functions (Noller and Callan, 1986; Feldman and Gehring, 1988; Carlson et al., 1991; Grych et al., 1992; Ohannessian et al., 2000). So they do not necessarily have the same idea of family harmony. Feldman and Gehring (1988) pointed out that "family interactions from children's statements and objective assessments are independent. They may be overlapped, but they are both worth studying." From this point of view, the individual perception and evaluation of family harmony are different from the relatively objective comprehensive evaluation of family harmony. Experience of family harmony is the individual's own subjective evaluation of whether the family relationship is harmonious or not.

Through searching the literature, we can find that there is no measuring tool for family harmony in foreign countries because family harmony is a special topic based on Chinese cultural characteristics, while it is more commonly referred to as healthy family, strong families, happy families, stable families, successful families, optimal families, well-functioning families, and so on in the context of western culture (Wolcott, 1999; Siu and Shek, 2005; Ip, 2014; Fauziah, 2020). Although, in essence, these families also aim to achieve the goal of family harmony and happiness, there are still differences in specific standards and denotation between these families and harmonious families (Yang and Liu, 2008).

The family-related measurement tools developed abroad mainly include Family APGAR Questionnaire (Smilkstein, 1978), Family Environment Scale (Moos and Moos, 1981), Family Adaptability and Cohesion Evaluation Scale (FACES; Olson et al., 1982), Family Assessment Device (FAD; Kabacoff et al., 1990), Self-Report Family Inventory (Beavers et al., 1985). Some domestic scholars have also developed Chinese Family Assessment Instrument (Siu and Shek, 2005; Shek and Ma, 2010; Mellor et al., 2014), Relationship-Specific Chinese Family Assessment Instrument (Liu et al., 2011). Although these measurement tools are related to the content of family harmony, they mainly focus on the family function, and the emphasis is not family harmony.

At present, the main measuring tools for family harmony in China are as follows: (1) Harmonious family evaluation system constructed by evaluating 1,200 families in urban and rural areas of Shanghai (Xu, 2009); (2) Family Harmony Scale and the corresponding five-item simplified version scale, FHS-5 (Kavikondala et al., 2016); (3) Family harmony scale for adolescents (Li, 2016). However, there are also limitations in using the above scales to measure college students' experience of family harmony. First of all, the measurement of family harmony of the first two is a comprehensive assessment of family harmony by different members of the family (including husband, wife, father, mother, children, grandparents, siblings, etc.). Fang et al. (2004) pointed out that most previous studies require a member of the family to evaluate the family function. In fact, this is based on the premise that all family members have the same perception of the family function. Therefore, this kind of evaluation is not necessarily in line with reality, and the evaluation results are general and depersonalized. In addition, although the latter is developed for teenagers in the family, it similarly emphasizes the general form of family interaction. There is still a lack of description of the perspective of individual experience. For example, the items in the questionnaire such as "good health of family members, no major diseases" and "high quality of elders in the family" still focus on the relatively objective description of the family like many other similar family assessment instruments (Shek and Ma, 2010), rather than the subjective harmonious feeling and experience of the individual. Therefore, none of the above scales can accurately evaluate the college students' experience of family harmony. A person's cognitive assessment and perception of the meaning of life events are very important, and relevant studies have shown that external support can have a positive impact on the individual only after it is perceived and recognized by the individual (Rutter, 1981; Li and Yin, 2015). Therefore, the experience of family harmony, as the perception of whether the family is harmonious or not, is more important for the development of his or her physical and mental health. However, previous studies have not dealt with the theme of college students' experience of family harmony. On the one hand, the development of measuring tools for college students' family harmony can enrich relevant research in the field of family harmony and help college students to further understand their cognition of family relations. On the other hand, they can understand their experiences of family relationship. And this study can provide more reference information for family therapy and carry out more targeted work. Therefore, the purpose of the current study was to develop an instrument for measuring the family harmony experience of college students in China on the basis of the existing research and test it from the perspective of psychometrics.

METHOD

Participants

Sample 1: Here, 600 college students from a University in Tianjin were selected as subjects. From them, 562 valid questionnaires were collected, with an effective recovery rate of 94%. Among the valid samples, there were 222 freshmen, 119 sophomores, 116 juniors, and 105 seniors (including 309 males and 253 females). The ages of the students ranged from 17 to 23 ($M = 19.35$, $SD = 1.39$). Sample 1 is used for item analysis and exploratory factor analysis (EFA) of the initial questionnaire.

Sample 2: Here, 800 college students from a University in Tianjin were selected, and 696 valid questionnaires were obtained. The effective recovery rate was 87%. Among the valid samples, there were 271 freshmen, 169 sophomores, 139 juniors, and 117 seniors (including 350 males and 346 females). The ages of the students ranged from 17 to 23 ($M = 19.30$, $SD = 1.35$). Sample 2 was used for confirmatory factor analysis (CFA), construct validity analysis, and internal consistency reliability analysis of the formal questionnaire.

Sample 3: Here, 618 college students from a University in Tianjin were selected, and 519 valid questionnaires were obtained. The effective recovery rate was 83.98%. Among the valid samples, there were 230 freshmen, 127 sophomores, 98 juniors, and 64 seniors (including 280 males and 239 females). The ages of the students ranged from 16 to 23 ($M = 19.15$, $SD = 1.27$). Sample 3 was used to analyze the criterion-related validity of the formal questionnaire.

Sample 4: Here, 248 college students were selected from sample 3 and retested 2 months later. There were 223 valid questionnaires, with an effective recovery rate of 89.92%. Among the valid samples, there were 154 freshmen, 50 sophomores, 11 juniors, and eight seniors (including 100 males and 123 females). This group ranged from 17 to 23 years of age ($M = 18.96$, $SD = 1.03$). Sample 4 was used to analyze the test–retest reliability of the formal questionnaire.

Chinese families are traditionally all heterosexual families, and there are no homosexual families by law. So all of the college students were from heterosexual families. Among them, 618 college students were investigated in this study, among which 35 were from single-parent families, 577 were from non-single-parent families, and six were missing. And there was no significant difference in the total score of harmony between single-parent and non-single-parent families ($p > 0.05$).

Procedure of Development of the Questionnaire

Construction of Dimensions

We have drawn up an open-ended questionnaire on the experience of family harmony, which requires college students to write three items of the experience of family harmony and family disharmony according to their own actual situation. Here, 327 subjects were distributed in a University in Tianjin, including 171 males and 156 females. All the items were collected and sorted out to be 981 items of the experience of family harmony and 976 items of the experience of family disharmony, and the survey results were classified and sorted out. Then, combined with the relevant theories and measurements at home and abroad, it initially formed five aspects: Atmosphere of family, Responsibility to housework, Time-sharing, Seeking help, Supporting family members (ARTSS). Considering the multi-perspective evaluation of college students' experience of family harmony, each aspect consists of the opposite bipolar perspectives to construct specific dimensions and finally form 10 subdimensions. They include getting along vs. conflict, undertaking housework vs. refusing housework, sharing vs. self-isolatedness, help-seeking vs. avoidance, and support-providing vs. indifference, respectively.

Preparation of the Items

First of all, the representative items were selected according to the open-ended questionnaire. For example, "There is always full of laughter among my family members at home" belongs to "Getting along." "My family members often quarrel with each other" belongs to "Conflict." "We share interesting stories together" belongs to "Sharing." "We don't have enough time to get along and communicate with each other" belongs to "Self-isolatedness." "I can help my family members when they are in trouble" belongs to "Support-providing." "My family and I don't care about each other" belongs to "Indifference." "I communicate with my family members as soon as possible when something happens to me" belongs to "Help-seeking." "There's no one to talk about my pain at home" belongs to "Avoidance." "We don't evade the responsibility of housework" belongs to "Undertaking housework." A total of 29 items were extracted from the open-ended questionnaire of college students.

Secondly, some items originated from Family Environment Scale, FACES, FAD, The Chinese version of FAD, Family APGAR Questionnaire, and Family Harmony Scale. For example, after referring to the item "I am satisfied with the amount of time my family and I spend together" in the Family APGAR Questionnaire, we modified this item as "I am very satisfied that my family members spend time with me." Referring to the items in the Family Harmony Scale, "Family members listen to each other's opinions," we modified this item as "We will listen to each other's opinions when we meet problems" and so on. There are 34 items extracted from the above six questionnaires.

Thirdly, an expert group of four psychologists is formed to discuss whether each item accurately expresses the meaning represented by the corresponding dimensions and whether the expression is appropriate, so as to make appropriate deletions and modifications. For example, the objects of the questionnaire are the college students in a family, not all family members, so the description about the subjects of some items in the existing scale has been modified, such as "family members take the initiative to talk to family members" being modified as "I will take the initiative to talk to family members" and "Other people will pay attention when family members encounter troubles" being modified as "I will pay attention to family members when they encounter troubles." Finally, 63 items about the experience of family harmony were obtained, and the initial questionnaire was formed. Among them, there are nine items of getting along, seven items of conflict, nine items of undertaking housework, three items of refusing housework, eight items of sharing, five items of self-isolatedness, six items of help-seeking, six items of avoidance, seven items of support-providing, and three items of indifference. The dimensions of the initial questionnaire and the specific sources of all items are shown in **Table 1**.

TABLE 1 | Dimensions and items of the initial questionnaire for college students' experience of family harmony.

Dimension	Item	Source of the item
Getting along	Q1. Every member in my family is free to express his/her opinions.	3
	Q20. My family members can be modest to each other when there is a conflict in the family.	3
	Q21. My family members always get along with each other.	2
	Q38. I feel that everyone in my family is backing each other up.	1
	Q39. My family members don't have to be careful when they communicate with each other.	Modified from 1
	Q53. I don't feel stressed at home.	Modified from 1
	Q54. My family members love each other.	1
	Q61. There is always full of laughter among my family members at home.	1
	Q63. We seldom have family conflicts.	1
Conflict	Q2. I feel like the atmosphere at home is depressing and suffocating.	1
	Q19. My family members have a cold war with each other.	1
	Q22. My family members often quarrel with each other.	1
	Q37.I feel like not to stay at home.	1
	Q40. My family members complain about each other when things go wrong.	1
	Q52. My family members are seldom gentle and considerate to each other.	5
	Q55. My family members often blame and criticize each other.	2
Undertaking housework	Q9. Do it together if something needs to be dealt with at home	3
	Q12. We can share the housework together.	Modified from 4
	Q29. We will discuss the division of housework.	Modified from 4
	Q31. We take turns to share different housework in the family.	3
	Q46.We all share family obligations.	Modified from 3
	Q47. We are willing to spend a lot of energy doing things at home.	Modified from 2
	Q58. We don't evade the responsibility of housework.	1
	Q59. Everyone in my family does his/her job.	1
	Q62. We do housework together.	1
Refusing housework	Q10. We complain to each other that the other side did too little housework.	Modified from 4
	Q11. The housework of our family focuses on individual people.	5
	Q30. Few people volunteer to do something at home.	2
Sharing	Q7. We participate in things we are all interested in.	4
	Q14. We will discuss and consult together when we encounter problems.	Modified from 4
	Q27. We'll show each other our love.	Modified from 4
	Q33. I am very satisfied that my family members spend time with me.	Modified from 6
	Q44. We will listen to each other's opinions when we meet problems.	Modified from 7
	Q48. I will try my best to spend time with my family members.	1
	Q57. My family members take part in recreational activities together.	Adapted from 3
	Q60. We share interesting stories together.	1
Self-isolatedness	Q8. We don't express our love for each other.	5
	Q13. We prefer to do things separately rather than with the whole family.	3
	Q28. I seldom consider other family members' opinions when I do things.	Modified from 2
	Q32.We don't have enough time to get along and communicate with each other.	Modified from 1
	Q45. There is little time for my family members to spend time with each other.	1
Help-seeking	Q3. I can get comfort and help at home when I encounter difficulties.	1
	Q18. I will take the initiative to talk to my family.	Modified from 3
	Q23. I can tell my family about my difficulties and troubles.	Modified from 2
	Q36. I communicate with my family members as soon as possible when something happens to me.	1
	Q51. I will discuss the solution with my family if I have a problem.	Modified from 4
Avoidance	Q4. There's no one to talk to about my pain at home.	1
	Q17. It's hard to talk to my family when I come across something that makes me sad.	Modified from 5
	Q24.I don't talk to my family when I'm angry.	5

(Continued)

TABLE 1 | Continued

Dimension	Item	Source of the item
	Q35. I choose to take it alone when I have something to worry about.	1
	Q42. I never tell my family what's on my mind.	4
	Q50.I don't tell my family what happened.	2
Support-providing	Q5.I will pay attention to my family members when they are in trouble.	Modified from 4
	Q16. We can support each other in times of crisis.	Modified from 4
	Q25. My family can accept and support it when I engage in new activities.	Modified from 6
	Q34. I will support the ideas or decisions of other family members.	1
	Q41. My family would like to listen to my opinions and ideas patiently and support me as much as possible.	1
	Q43. I can help my family members when they are in trouble.	Modified from 1
	Q49. I will care for my family members.	Modified from 1
	Q56. I can give warmth and comfort to my family members when they need it.	Modified from 1
Indifference	Q6. I'm self-centered and I don't care about my family.	Modified from 4
	Q15. My family and I don't care about each other.	Modified from 1
	Q26. My family members only care about themselves and ignore the family.	Modified from 1

1, Open-ended survey of college students; 2, Family Environment Scale; 3, Family Adaptability and Cohesion Evaluation Scale; 4, Family Assessment Device; 5, the Chinese version of FAD; 6, Family APGAR Questionnaire; 7, Family Harmony Scale.

Formation of the Questionnaire

The subjects of sample 1 were measured with the initial questionnaire of college students' experience of family harmony. In the item analysis, one item with poor differentiation is deleted according to the standard that the correlation coefficient between each item and the total score is <0.4 (Wu, 2010). The measurement results of the remaining 62 items were analyzed with EFA. Such items would be deleted if factor loadings were <0.4, the number of the items in a factor was <3, an item has excessive loading on multiple factors, and having a discrepancy of meaning between the item and the related factor (Wu, 2010). Finally, a formal questionnaire was formed, including 56 items in nine dimensions. The subjects of sample 2 were measured with the formal questionnaire. CFA, construct validity analysis, and internal consistency reliability analysis were carried out, and criterion-related validity was analyzed after measurement of sample 3. The participants in sample 4 were retested 2 months later for test–retest reliability analysis (see the Results section for details).

Instruments

Open-Ended Questionnaire for College Students

We have prepared an open-ended questionnaire for college students' experience of family harmony, which contains two questions: (1) How do you experience family harmony? Please write at least three items. (2) How do you experience family disharmony? Please write at least three items.

Initial Questionnaire on College Students' Experience of Family Harmony

The self-designed initial questionnaire contains five groups of evaluation indicators with bipolarity, namely, a total of 10 dimensions, which are getting along, conflict, undertaking housework, refusing housework, sharing, self-isolatedness, help-seeking, avoidance, support-providing, and indifference. The questionnaire consists of 63 items, including 24 reverse scoring items. Scoring of each item in the College Students' Experience of Family Harmony Questionnaire (CSEFHQ) is based on a 4-point Likert scale, which ranges from 1 to 4 (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree), and a reverse score was given to 24 items reflecting experience of family disharmony.

Formal Questionnaire on College Students' Experience of Family Harmony

The self-designed formal questionnaire consists of nine dimensions, including getting along, conflict, undertaking housework, sharing, self-isolatedness, help-seeking, avoidance, support-providing, and indifference. This questionnaire has 56 items, each with the same scoring method as the initial questionnaire, including 21 reverse scoring items. The higher the final score is, the higher the experience of family harmony is.

Criterion Questionnaire

(1) **Family APGAR Questionnaire**

The Family APGAR Questionnaire was compiled by Smilkstein (1978) to evaluate Adaption, Partnership, Growth, Affection, and Resolve of family function. The questionnaire has five items. Respondents use a 3-point Likert-type scale (ranging from *never* to *often*) for each item. The higher the score is, the higher the family support is. In this study, the Cronbach's alpha reliability of the questionnaire is 0.86.

(2) **Family Adaptability and Cohesion Evaluation Scale**

The FACES II was compiled by Olson et al. (1982). Fei et al. (1991) translated and revised the FACES II-CV. There were 30 items in the evaluation of family intimacy and family adaptability. Respondents use a 5-point Likert-type scale (ranging from *never* to *always*) with each item. In this study, the Cronbach's alpha reliability of family intimacy, family adaptability, and the total scale was 0.88, 0.88, 0.94, respectively.

(3) **Well-Being Scale**

The Index of Well-Being (IWB) Scale, compiled by Campbell (1976), is used to measure the degree of happiness experienced by individuals at present, including the overall emotion index and life satisfaction index, a total of nine items. Scores on this scale are based on nine items with a 7-point Likert scale. The higher the score is, and the higher the happiness is. In this study, the Cronbach's alpha reliability of the scale is 0.91.

(4) **Loneliness Scale**

The UCLA Loneliness Scale, compiled by Russel et al. (1978), is used to evaluate individual loneliness caused by the gap between the desire for social communication and the actual level. Scores on this scale are based on 20 items with a 4-point Likert scale ranging from *never* to *often*. The higher the total score is, the higher the loneliness is. In this study, the Cronbach's alpha reliability of the scale is 0.91.

Data Processing

After the questionnaire was collected uniformly, the data were input into Statistical Package for the Social Sciences (SPSS) software. The researchers used SPSS 21.0 and AMOS 22.0 to analyze the data.

RESULTS

Item Analysis of the Initial Questionnaire

Two methods are used in the item analysis. The first is the critical ratio method, in which the total scores of the items are sorted from high to low, and the samples are divided into the top 27% and the last 27%, to test whether there is a significant difference between the score of top-score group and that of low-score group on each item, that is, the significance of the item critical value (CR). The results show that the *p*-value of each item is <0.001, and the discrimination degree is ideal. The second is the correlation method. The Pearson correlation coefficient (r) between the scores of each item and the total score of family harmony shows that the 10th item should be deleted because the correlation coefficient is <0.4 (Wu, 2010). However, the correlation coefficients of the remaining 62 items are between 0.40 and 0.75, and all reach a significant level of 0.01. There is a good item differentiation.

Exploratory Factor Analysis of the Initial Questionnaire

The remaining 62 items in the initial questionnaire were analyzed with EFA. The results showed that the Kaiser–Meyer–Olkin value is 0.963, exceeding the recommended value of 0.6, as suggested by Kaiser (1974). Bartlett's Test of Sphericity (Bartlett, 1950) is also significant ($\chi 2 = 23939.096$, df $= 1891$, $p < 0.001$), which indicated that the set of correlations in the correlation matrix were significantly different from zero and suitable for factor analysis. The principal component analysis with variance maximum rotation method is used to extract common factors, and the number of factors is determined according to the Eigenvalue exceeding 1 (Kim and Mueller, 1978). Deleted items are with reference to the following criteria: (1) factor loading is <0.4; (2) commonality is <0.2; (3) the factor contains <3 items; (4) the loading on multiple factors is too high; (5) the meaning of the item is not consistent with the meaning of the factor (Wu, 2010). After deleting six items (items 9, 11, 25, 30, 41, and 58), the CSEFHQ resulted in the extraction of nine factors consisting of 56 items, and the cumulative variance explanation rate is 66.03%. The factor loading of each item is shown in **Table 2**.

As shown in **Table 2**, F1 consists of nine items, which belong to the getting along in the theoretical conception; F2 consists of seven items, which belong to the conflict in the theoretical conception; F3 consists of seven items, which belong to the undertaking housework in the theoretical conception; F4 consists of six items, which belong to the avoidance in the theoretical conception; F5 consists of six items, which belong to the support-providing in the theoretical conception. F6 consists of eight items, which belong to the sharing in the theoretical conception; F7 consists of five items, which belong to the help-seeking in the theoretical conception; F8 consists of five items, which belong to the self-isolatedness in the theoretical conception; and F9 consists of three items, which belong to the indifference in the theoretical conception. There is a total of 56 items, including 21 reverse scoring questions.

Confirmatory Factor of Analysis of the Formal Questionnaire

According to the results of EFA, we decided to run confirmatory factor analysis of the nine factors that consist of 56 items. The CFA was performed using the AMOS 22.0 program. The following commonly used criteria were utilized in evaluating the adequacy of the model: the value of $\chi^2/df < 3$ is good. Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), and Incremental Fit Index (IFI) have values ranging from 0 to 1, and the values closer to 1 are much better. The values of root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR) that are <0.08 indicate that the model fits well on the whole (Wen et al., 2004). The results of this study showed that $\chi^2/df = 2.71 < 3$, $p < 0.001$. And the values of *IFI*, *TLI*, and *CFI* were all >0.85. $RMSEA = 0.05 < 0.08$, $SRMR = 0.05 < 0.08$. Overall, the main fit indexes for the nine-factor model were acceptable (**Table 3**), indicating that the formal questionnaire structure meets the expectation. In the path diagram of CFA of this model, the standardized factor loading of each item on the factors is >0.5 and significant (**Figure 1**). These results show that the formal questionnaire has good construct validity.

Correlation Analysis Among the Scores of Each Factor of the Formal Questionnaire

In order to further test the structural validity of the formal questionnaire, the correlations among the factors of the CSEFHQ and those between the factors and the total score were analyzed. The correlation coefficient between each factor is at a medium

TABLE 2 | Nine factors of the CSEFHQ and their factor loadings ($N = 562$).

Factors	Items	Loadings
F1: Getting along	Q53. I don't feel stressed at home.	0.69
	Q39. My family members don't have to be careful when they communicate with each other.	0.68
	Q54. My family members love each other.	0.68
	Q63. We seldom have family conflicts.	0.67
	Q21. My family members always get along with each other.	0.65
	Q20. My family members can be modest to each other when there is a conflict in the family.	0.63
	Q38. I feel that everyone in my family is backing each other up.	0.63
	Q61. There is always full of laughter among my family members at home.	0.62
	Q1. Every member in my family is free to express his/her opinions.	0.61
F2: Conflict	Q19. My family members have a cold war with each other▲.	0.75
	Q40. My family members complain about each other when things go wrong▲.	0.73
	Q55. My family members often blame and criticize each other▲.	0.72
	Q22. My family members often quarrel with each other▲.	0.71
	Q52. My family members are seldom gentle and considerate to each other▲.	0.70
	Q37.I feel like not to stay at home▲.	0.69
	Q2. I feel like the atmosphere at home is depressing and suffocating▲.	0.63
F3: Undertaking housework	Q31. We take turns to share different housework in the family.	0.79
	Q29. We will discuss the division of housework.	0.78
	Q12. We can share the housework together.	0.75
	Q62. We do housework together.	0.74
	Q46. We all share family obligations.	0.67
	Q59. Everyone in my family does his/her job.	0.54
	Q47. We are willing to spend a lot of energy doing things at home.	0.53
F4: Avoidance	Q35. I choose to take it alone when I have something to worry about▲.	0.76
	Q42. I never tell my family what's on my mind▲.	0.75
	Q24. I don't talk to my family when I'm angry▲.	0.71
	Q17. It's hard to talk to my family when I come across something that makes me sad▲.	0.69
	Q50. I don't tell my family what happened▲.	0.67
	Q4. There's no one to talk about my pain at home▲.	0.66
F5: Support-providing	Q49. I will care for my family members.	0.72
	Q5. I will pay attention to my family members when they are in trouble.	0.70
	Q43. I can help my family members when they are in trouble.	0.70
	Q56. I can give warmth and comfort to my family members when they need it.	0.64
	Q16. We can support each other in times of crisis.	0.63
	Q34. I will support the ideas or decisions of other family members.	0.60
F6: Sharing	Q44. We will listen to each other's opinions when we meet problems.	0.58
	Q33. I am very satisfied that my family members spend time with me.	0.58
	Q48. I will try my best to spend time with my family members.	0.58
	Q60. We share interesting stories together.	0.57
	Q57. My family members take part in recreational activities together.	0.55
	Q27. We'll show each other our love.	0.54
	Q14. We will discuss and consult together when we encounter problems.	0.52
	Q7. We participate in things we are all interested in.	0.45
F7: Help-seeking	Q23. I can tell my family about my difficulties and troubles.	0.72
	Q18. I will take the initiative to talk to my family.	0.71
	Q36. I communicate with my family members as soon as possible when something happens to me.	0.67
	Q3. I can get comfort and help at home when I encounter difficulties.	0.65
	Q51. I will discuss the solution with my family if I have a problem.	0.64
F8: Self-isolatedness	Q32. We don't have enough time to get along and communicate with each other▲.	0.75
	Q45. There is little time for my family members to spend time with each other▲.	0.65

(Continued)

TABLE 2 | Continued

Factors	Items	Loadings
	Q28. I seldom consider other family members' opinions when I do things▲.	0.60
	Q13. We prefer to do things separately rather than with the whole family▲.	0.59
	Q8. We don't express our love for each other▲.	0.59
F9: Indifference	Q15. My family and I don't care about each other▲.	0.80
	Q6. I'm self-centered and I don't care about my family▲.	0.67
	Q26. My family members only care about themselves and ignore the family▲.	0.59

CSEFHQ, College Students' Experience of Family Harmony Questionnaire. F1–F9 represent the nine extracted factors. ▲ Indicates negative scoring items. Item number originated from initial questionnaire.

TABLE 3 | Fit indexes of CFA of the CSEFHQ ($N = 696$).

$\chi 2$	*df*	$\chi 2/df$	IFI	TLI	CFI	RMSEA	SRMR
3,918.99	1,448	2.71	0.90	0.89	0.90	0.05	0.05

CFA, confirmatory factor analysis; CFI, Comparative Fit Index; CSEFHQ, College Students' Experience of Family Harmony Questionnaire; IFI, Incremental Fit Index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; TLI, Tucker–Lewis Index.

or low level, indicating that each factor has the same direction, but there are certain differences. The correlation coefficient between each factor and the total score is at a medium or high level, indicating that each factor has a good consistency with the total questionnaire. The results showed that the correlation coefficients among the factors were between 0.37 and 0.79, and the correlation was significant ($p < 0.01$). The correlation coefficient between each factor and the total score was between 0.61 and 0.91, and the correlation was significant ($p < 0.01$). The correlation between each factor and the total score was higher than that between each factor (**Table 4**). In summary, the questionnaire has good structural validity.

Criterion Validity Analysis of the Formal Questionnaire

The index of family care reflects the subjective satisfaction with the family, and family intimacy is the evaluation of the emotional connection and intimate relationship of family members. In addition, studies have shown that family relationships are closely related to individual happiness and loneliness (Duan, 1996; Deng and Zheng, 2013). Therefore, the Family APGAR Questionnaire, FACES, the IWB scale, and UCLA Loneliness Scale were selected as criterion tools. The results showed that the scores of each factor and the total scores of the CSEFHQ are correlated significantly with measures of APGAR ($r = 0.73$, $p < 0.01$), cohesion ($r = 0.71$, $p < 0.01$), adaptability ($r = 0.75$, $p < 0.01$), well-being ($r = 0.51$, $p < 0.01$), and loneliness ($r = -0.56$, $p < 0.01$). It shows that the criterion validity of the CSEFHQ is good (**Table 5**).

Reliability Analysis of the Formal Questionnaire

This study examined the internal consistency reliability, split-half reliability, and test–retest reliability of the formal questionnaire. The results show that the Cronbach's alpha coefficient of the total questionnaire is 0.97, and the Cronbach's alpha coefficient of each factor is 0.66–0.91. All these coefficients are >0.60, indicating that the questionnaire has good internal consistency and the reliability was at a good level (Wu, 2009). The split-half reliability of the total questionnaire is 0.92, and the split-half reliability of each factor is 0.67–0.90, indicating that the questionnaire has good equivalence. The test–retest reliability of the total questionnaire is 0.75, and the test–retest reliability of each factor is 0.49–0.72, and correlations are significant. According to relevant criteria, the test–retest reliability is acceptable (Robinson et al., 2010), indicating that the questionnaire has good cross-time stability. Overall, the CSEFHQ has good reliability, and the specific reliability indicators of the questionnaire are shown in **Table 6**.

DISCUSSION

On the basis of literature analysis and survey with questionnaires and reference to the existing research at home and abroad, this study first constructed five pairs of evaluation indicators of the college students' experience of family harmony with bipolarity. The reason for evaluating college students' experience of family harmony from the perspective of positive and negative experience is that two opposite perspectives can lead to a more real family harmony experience. Thus, it makes the evaluation results more accurate and reliable. At the same time, it can more comprehensively and accurately reflect all aspects of college students' experience of family harmony. Secondly, the initial questionnaire is formed by establishing the questionnaire items through the open-ended questionnaire, reference to the relevant scale, and expert group evaluation. Then, the development of the questionnaire has been finally completed through the pretest, formal test, and retest to test the reliability and validity of the questionnaire.

The developmental process of the whole questionnaire is as follows:

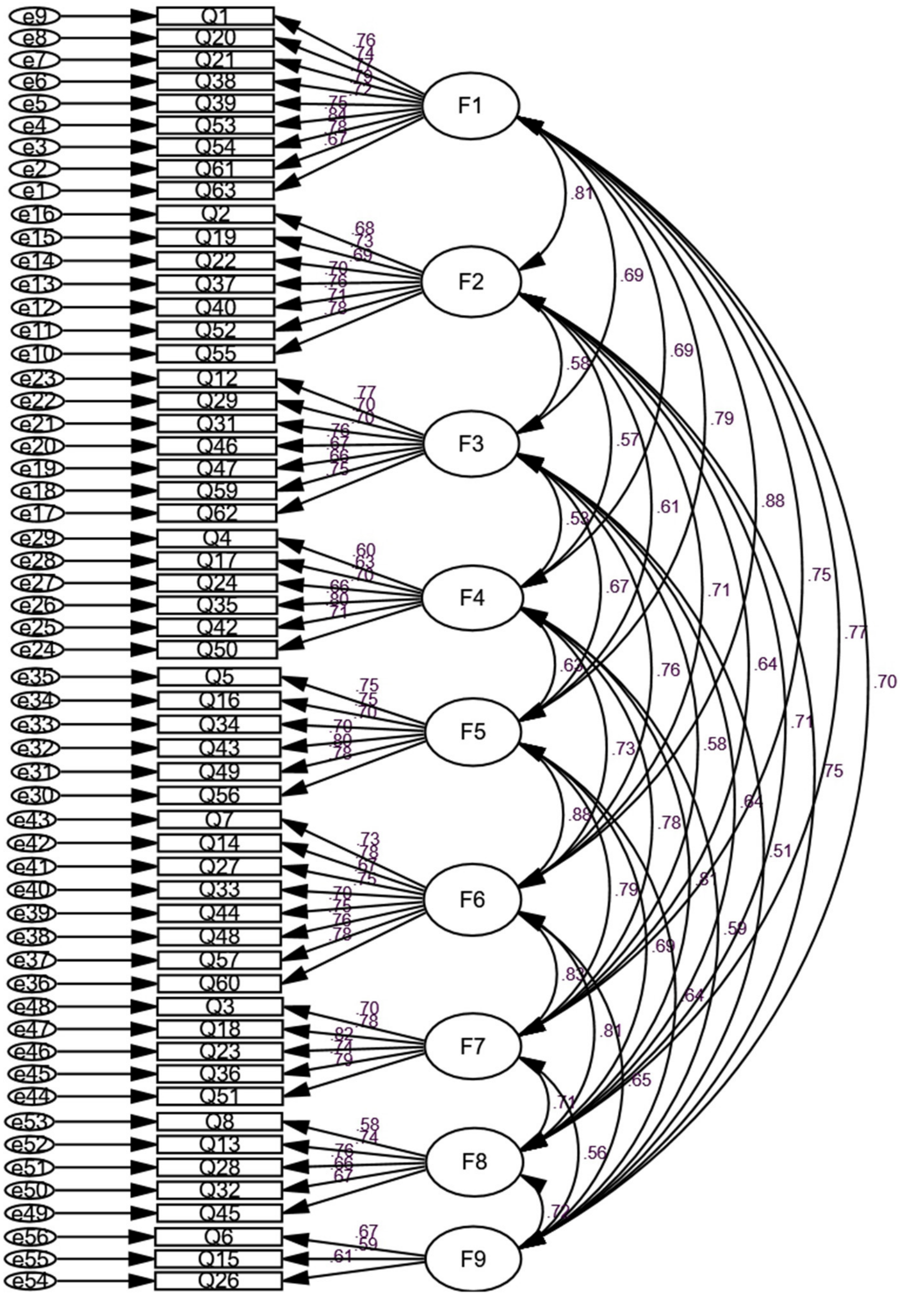

FIGURE 1 | Path diagram and estimated parameter loadings for the nine-factor model of the college students' experience of family harmony questionnaire (CSEFHQ) (N = 696). Latent variables (factors F1–F9) are indicated with ovals, and observed variables (items Q1–Q63) are indicated with rectangles.

The initial questionnaire consisted of 63 items with 10 dimensions (or factors): getting along, conflict, undertaking housework, refusing housework, sharing, self-isolatedness, help-seeking, avoidance, support-providing, and indifference. According to the results of item analysis, the 10th item Q10 (we complain to each other that the other side did too little correlation) should be deleted because the correlation coefficient is <0.4. Then, EFA was performed, and six items have been deleted. The 11th item Q11 (the housework of our family focuses on individual people) and the 41st item Q41 (my family would like to listen to my opinions and ideas patiently and support me as much as possible) were deleted due to multiple loads. The ninth item Q9 (do it together if something needs to be dealt with at home), 25th item Q25 (my family can accept and support it when I engage in new activities), and 58th item Q58 (we don't evade the responsibility of housework) have been deleted according to matching of the item meaning. The 30th item Q30 (few people volunteer to do something at home) has

TABLE 4 | The correlation between the scores of factors and the total score of the CSEFHQ ($N = 696$).

	F1	F2	F3	F4	F5	F6	F7	F8	F9
F2	0.72**								
F3	0.62**	0.50**							
F4	0.61**	0.50**	0.46**						
F5	0.69**	0.53**	0.59**	0.55**					
F6	0.79**	0.63**	0.67**	0.64**	0.78**				
F7	0.67**	0.56**	0.51**	0.67**	0.69**	0.74**			
F8	0.64**	0.58**	0.54**	0.66**	0.57**	0.69**	0.60**		
F9	0.54**	0.57**	0.37**	0.44**	0.48**	0.49**	0.42**	0.50**	
CSEFHQ	0.89**	0.77**	0.75**	0.77**	0.81**	0.91**	0.82**	0.80**	0.61**

$^{**}p < 0.01$. CSEFHQ, College Students' Experience of Family Harmony Questionnaire.

TABLE 5 | Correlation coefficients between factors and criteria of the CSEFHQ ($N = 519$).

	F1	F2	F3	F4	F5	F6	F7	F8	F9	CSEFHQ
APGAR	0.66**	0.61**	0.48**	0.52**	0.58**	0.64**	0.68**	0.53**	0.40**	0.73**
Cohesion	0.64**	0.49**	0.59**	0.56**	0.48**	0.70**	0.61**	0.52**	0.35**	0.71**
Adaptability	0.70**	0.56**	0.64**	0.56**	0.52**	0.70**	0.63**	0.53**	0.36**	0.75**
Well-being	0.44**	0.39**	0.36**	0.37**	0.44**	0.45**	0.44**	0.40**	0.22**	0.51**
Loneliness	−0.44**	−0.45**	−0.44**	−0.40**	−0.52**	−0.49**	−0.48**	−0.45**	−0.30**	−0.56**

$^{**}p < 0.01$. CSEFHQ, College Students' Experience of Family Harmony Questionnaire.

been deleted because it is inconsistent with the meaning of the factor "self-isolatedness." The questionnaire remains to have 56 items after seven items were deleted. Since all three items in the factor "refusing housework" had been deleted (Q10, Q11, and Q30), the final formal questionnaire only included nine factors: getting along, conflict, undertaking housework, sharing, self-isolatedness, help-seeking, avoidance, support-providing, and indifference.

Among the nine extracted factors, the two factors of "getting along" and "conflict" explain the highest ratio of variance in EFA. In addition, by conducting a survey with the open-ended questionnaires, it is also found that the highest proportions of family harmony and family disharmony experienced by college students are "family members getting along well" and "family members often have conflicts and disputes," respectively. This shows that the harmonious coexistence of family members is the most important part of college students' experience of family harmony. Shek and Man-fei (2000) believe that the ideal family in Chinese Confucianism puts great emphasis on family harmony and avoids quarrels. A survey conducted by Shek (2001) in Hong Kong also shows that people place special emphasis on avoiding conflicts in family harmony. Referring to the existing measurement tools related to family harmony (Xu, 2009; Kavikondala et al., 2016; Li, 2016), the dimension of housework responsibility is not involved, so this is a significant feature of the CSEFHQ. Housework is an important part of family life. Family harmony not only is the close and harmonious relationship between family members but also includes the responsibility for family affairs. Family members' division of labor and shared family obligations are one of the elements of family happiness and harmony (Wolcott, 1999; Fauziah, 2020). If housework is concentrated on individual members, it is easy to cause complaints, dissatisfaction, and even disputes. The study of Liu et al. (2015) found that housework is a kind of emotional work and is not simple labor. Sharing housework contributes to the intimacy and harmony of family members, especially husband and wife. Therefore, housework responsibility is of great significance to family harmony, and it is also an important part of college students' experience of family harmony. Help-seeking and avoidance and support and indifference reflect whether college students can get help in their families and whether they can provide support to their families. These are the manifestations of interpersonal interaction among family members, which are similar to the mutuality dimension in the evaluation of Chinese family function by Siu and Shek (2005), emphasizing mutual support of family members. Sharing and self-isolatedness reflect whether college students spend time with their families and share happiness or less than companionship and lack of sharing. Previous studies have shown that family interaction following the sharing model contributes to family harmony (Chuang, 2005), and sharing family time is the core of family harmony and happiness (Lam et al., 2012). Therefore, sharing with family members is also an important part of college students' experience of family harmony. The experience

TABLE 6 | Reliability of the CSEFHQ (Internal consistency and split-half: $N = 696$; Test–retest: $N = 223$).

	F1	F2	F3	F4	F5	F6	F7	F8	F9	CSEFHQ
Internal consistency	0.91	0.88	0.88	0.84	0.88	0.90	0.87	0.81	0.66	0.97
Split-half	0.90	0.84	0.81	0.82	0.87	0.89	0.86	0.84	0.67	0.92
Test–retest	0.68**	0.63**	0.63**	0.72**	0.60**	0.68**	0.66**	0.62**	0.49**	0.75**

$^{**}p < 0.01$. CSEHFQ, College Students' Experience of Family Harmony Questionnaire.

of family harmony has an important influence on the physical and mental development of college students (Johnson et al., 2010; Lucas-Thompson and Hostinar, 2013; Rhoades and Wood, 2014; Cheung et al., 2019), but there is no measurement tool for this aspect at present. So, on the one hand, the development of the CSEFHQ can enrich relevant research in the field of family harmony. On the other hand, it can provide practical guidance for improving college students' experience of family harmony and better provide practical guidance for the development of college students' physical and mental health. However, due to the convenience of sampling, this study only selected the sample of college students in Tianjin but did not select more samples of national college students, so it needs to be further verified and improved in the future. Actually, that belongs to one limitation of this study. If this questionnaire is used in the general population, further research should be conducted. This study focuses on college students' experience of family harmony, and future research can further examine different family roles, such as college students' parents' experience of family harmony. By comparing the evaluation results, we can determine the possible differences of the experience of family harmony based on different perspectives and the influence of these differences on them. Finally, the questionnaire developed in this study is aimed at general and healthy families, and further research is needed if family members have mental illness.

CONCLUSION

This study developed an effective instrument to measure college students' experience of family harmony in China. The CSEFHQ with 56 items consists of nine factors: getting along and conflict, help-seeking and avoidance, support-providing and indifference, sharing and self-isolatedness, and undertaking housework. The structure of the questionnaire is reasonable, and the discrimination, reliability, and validity of the items meet the requirements of psychometrics. This study mainly discusses the experience of family harmony among Chinese college students, and the experience of family harmony among other groups besides college students can be further explored. In addition, due to the differences between different cultures, the experience of family harmony and its differences among different groups in the cross-cultural background can be further studied in the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee, Tianjin University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

QZ: construct the framework of the development of this questionnaire (CSEFHQ) and the related idea. QW: carried out this program into effect. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported in part by grants from the Tianjin Education Committee Scientific Research Program Special Task Key Project Research on competency model of psychological monitors implementing peer psychological service (grant number 2020ZXXL-GX19G) and Tianjin Education Committee Social Science Key Project Standardization construction and practice of mental health in colleges and universities (grant number 2020JWZD13).

ACKNOWLEDGMENTS

This study was the product of a master's dissertation conducted at Tianjin University. The authors thank all the experts and postgraduates participating in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.658430/full#supplementary-material

REFERENCES

Bartlett, M. (1950). Tests of significance in factor analysis. *Br. J. Stat. Psychol.* 3, 77–85. doi: 10.1111/j.2044-8317.1950.tb00285.x

Beavers, W. R., Hampson, R. B., and Hulgus, Y. F. (1985). Commentary: the beavers systems approach to family assessment. *Fam. Process*. 24, 398–405. doi: 10.1111/j.1545-5300.1985.00398.x

Campbell, A. (1976). Subjective measures of well-being. *Am. Psychol.* 31, 117–124. doi: 10.1037/0003-066X.31.2.117

Carlson, C. I., Cooper, C. R., and Spradling, V. Y. (1991). Developmental implications of shared versus distinct perceptions of the family in early adolescence. *New Dir. Child Dev*. 1, 13–32. doi: 10.1002/cd.23219915103

Cheung, R. Y. M., Leung, M. C., Chiu, H. T., Kwan, J. L. Y., Yee, L. T. S., and Hou, W. K. (2019). Family functioning and psychological outcomes in emerging adulthood: savoring positive experiences as a mediating mechanism. *J. Soc. Pers. Relat*. 36, 2693–2713. doi: 10.1177/0265407518798499

Chuang, Y. (2005). Effects of interaction pattern on family harmony and well-being: test of interpersonal theory, relational-models theory, and confucian ethics. *Asian J. Soc. Psychol*. 8, 272–291. doi: 10.1111/j.1467-839X.2005.00174.x

Deng, L., and Zheng, R. (2013). The relationship between family function and emotional expressiveness and loneliness of college students. *Psychol. Behav. Res*. 11, 223–228. doi: 10.3969/j.issn.1672-0628.2013.02.014

Duan, J. (1996). Overview of subjective well-being. *Psychol. Dyn*. 4, 46–51.

Fang, X., Xu, J., Sun, L., and Zhang, J. (2004). Family functioning: theory, influencing factors and their relationship to adolescent social adjustment. *Adv. Psychol. Sci*. 12, 544–553. doi: 10.3969/j.issn.1671-3710.2004.04.009

Fauziah, N. (2020). The concept of family's harmony in multiple cultural settings, what about the family harmony with autism children in Indonesia? A literature study. *Fam. J.* 28, 1–6. doi: 10.1177/1066480720904027

Fei, L., Shen, Q., and Zheng, Y. (1991). The preliminary evaluation of family adaptability and cohesion evaluation scale and family environment scale: the control study of normal families and Schizophrenia families. *Chin. Ment. Health J*. 5, 198–202.

Feldman, S. S., and Gehring, T. M. (1988). Changing perceptions of family cohesion and power across adolescence. *Child Dev.* 59, 1034–1045. doi: 10.2307/1130269

Grych, J. H., Seid, M., and Fincham, F. D. (1992). Assessing marital conflict from the child's perspective: the children's perception of interparental conflict scale. *Child Dev*. 63, 558–572. doi: 10.2307/1131346

Hannum, J. W., and Dvorak, D. M. (2004). Effects of family conflict, divorce, and attachment patterns on the psychological distress and social adjustment of college freshmen. *J. Coll. Stud. Dev.* 45, 27–42. doi: 10.1353/csd.2004.0008

Ip, P. (2014). Harmony as happiness? Social harmony in two Chinese societies. *Soc. Indicat. Res.* 117, 719–741. doi: 10.1007/s11205-013-0395-7

Johnson, V. K., Gans, S. E., Kerr, S., and LaValle, W. (2010). Managing the transition to college: family functioning, emotion coping, and adjustment in emerging adulthood. *J. Coll. Stud. Dev.* 51, 607–621. doi: 10.1353/csd.2010.0022

Kabacoff, R. I., Miller, I. W., Bishop, D. S., Epstein, N. B., and Keitner, G. I. (1990). A psychometric study of the McMaster Family Assessment Device in psychiatric, medical, and nonclinical samples. *J. Fam. Psychol.* 3, 431–439. doi: 10.1037/h0080547

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika* 39, 31–36. doi: 10.1007/BF02291575

Kavikondala, S., Stewart, S. M., Ni, M. Y., Chan, B. H., Lee, P. H., Li, K. K., et al. (2016). Structure and validity of family harmony scale: an instrument for measuring harmony. *Psychol. Assess*. 28, 307–318. doi: 10.1037/pas0000131

Kim, J. O., and Mueller, C. W. (1978). *Factor Analysis: Statistical Methods and Practical Issues Sage University Paper Series on Quantitative Applications in the Social Sciences*. Beverly Hills, CA: Sage. doi: 10.4135/9781412984256

Lam, W. W. T., Fielding, R., McDowell, I., Johnston, J., Chan, S., Leung, G. M., et al. (2012). Perspectives on family health, happiness and harmony (3H) among Hong Kong Chinese people: a qualitative study. *Health Educ. Res*. 27, 767–779. doi: 10.1093/her/cys087

Li, D. (2016). *The Mediating Effect of Regulatory Emotional Self-Efficacy on Family Harmony and Resilience in Adolescent*. The Dissertation of Tianjin University, Tianjin University, Tianjin, China.

Li, Z., and Yin, X. (2015). How social support influences hope in college students: the mediating roles of self-esteem and self-efficacy. *Psychol. Dev. Educat*. 31, 610–617. doi: 10.5539/elt.v10n12p158

Liu, A., Zhuang, J., and Zhou, Y. (2015). What kind of men do housework: emotional expression, financial dependence or gender equality? *Cluster Women's Stud. Papers* 3, 20–28. doi: 10.3969/j.issn.1004-2563.2015.03.002

Liu, X., Zhang, J., and Yeh, C. L. (2011). Investigating the validity of a multirater assessment of family functioning in China. *Soc. Behav. Personal.* 39, 773–783. doi: 10.2224/sbp.2011.39.6.773

Lucas-Thompson, R. G., and Hostinar, C. E. (2013). Family income and appraisals of parental conflict as predictors of psychological adjustment and diurnal cortisol in emerging adulthood. *J. Fam. Psychol.* 27, 784–794. doi: 10.1037/a0034373

Mellor, D., Xu, X., Wong, J., and Richardson, B. (2014). The factor structure of the chinese family assessment instrument adapted for parent report. *Assessment* 21, 60–66. doi: 10.1177/1073191111425855

Moos, R., and Moos, B. (1981). *Family Environment Scale Manual*. Palo Alto CA: Consulting Psycholosists Press.

Noller, P., and Callan, V. J. (1986). Adolescent and parent perceptions of family cohesion and adaptability. *J. Adolesc*. 9, 97–106. doi: 10.1016/S0140-1971(86)80030-6

Ohannessian, C. M., Lerner, R. M., Lerner, J. V., and vonEye, A. (2000). Adolescent-parent discrepancies in perceptions of family functioning and early adolescent self-competence. *Int. J. Behav. Dev*. 24, 362–372. doi: 10.1080/01650250050118358

Olson, D., McCubbin,. H., Barnes, H., Larsen, A., Muxen, M., and Wilson, M. (1982). *Family Inventories: Inventories Used in a National Survey of Families Across the Family Life Cycle*. St. Paul MN: Family Social Science. University of Minnesota.

Rhoades, G. K., and Wood, L. F. (2014). Family conflict and college-student social adjustment: the mediating role of emotional distress about the family. *Couple Fam. Psychol.* 3, 156–164. doi: 10.1037/cfp0000024

Robinson, M., Stokes, K., Bilzon, J., Standage, M., Brown, P., and Thompson, D. (2010). Test-retest reliability of the Military Pre-training Questionnaire. *Occup. Med.* 60, 476–483. doi: 10.1093/occmed/kqq073

Russel, D., Peplau, L. A., and Freguson, M. L. (1978). Developing a measure of loneliness. *J. Pers. Assess.* 42, 290–294. doi: 10.1207/s15327752jpa4203_11

Rutter, M. (1981). Stress, coping and development: some issues and some questions. *J. Child Psychol. Psychiatry* 22, 323–356. doi: 10.1111/j.1469-7610.1981.tb00560.x

Shek, D. T. L. (2001). Chinese adolescents and their parents' views on a happy family: implications for family therapy. *Fam. Ther*. 28, 73–104.

Shek, D. T. L., and Ma, C. M. S. (2010). The Chinese Family Assessment Instrument (C-FAI): hierarchical confirmatory factor analyses and factorial invariance. *Res. Soc. Work Pract.* 20, 112–123. doi: 10.1177/1049731509355145

Shek, D. T. L., and Man-fei, L. (2000). Conceptions of an ideal family in confucian thoughts: implications for individual and family counseling (in Chinese). *Asian J. Counsell.* 7, 85–103.

Siu, A. M. H., and Shek, D. T. L. (2005). Psychometric properties of the Chinese Family Assessment Instrument in Chinese adolescents in Hong Kong. *Adolescence* 40, 817–830.

Smilkstein, G. (1978). The family APGAR: a proposal for a family function test and its use by physicians. *J. Fam. Pract*. 6, 1231–1239.

Wen, Z., Hou, J., and Marsh. (2004). Structural equation model testing: Cutoff criteria for goodness of fit indices and Chi-square test. *Acta Psychologica Sinica.* 36, 186–194.

Wolcott, I. (1999). Strong families and satisfying marriages. *Fam. Mat*. 4, 21–30.

Wu, M. (2009). *Structural Equation Modeling: Operations and Applications of AMOS*. Chongqing: Chongqing University Press.

Wu, M. (2010). *Practice of Questionnaire Statistical Analysis: SPSS Operation and Application*. Chongqing: Chongqing University Press.

Xu, A. (2009). The discussion of the index system of the harmonious family and its impact mechanism: the empirical research in Shanghai. *Jiangsu Soc. Sci.* 2, 88–97. doi: 10.3969/j.issn.1003-8671.2009.02.014

Yang, X., and Liu, C. (2008). The theory of current construction of harmonious family and the realization path. *Soc. Sci. Nan Jing* 9, 99–105. doi: 10.3969/j.issn.1001-8263.2008.09.015

Yap, P. M., and Tan, B. H. (2011). Families' experience of harmony and disharmony in systemic psychotherapy and its effects on family life. *J. Fam. Ther.* 33, 302–331. doi: 10.1111/j.1467-6427.2011.00543.x

Zhang, P. (2016). *The Double Mediating of Self-Compassion, Perceived Social Support Between Adult Attachment and Interpersonal Trust on Undergraduates*. The Dissertation of Harbin Normal University, Harbin Normal University, Harbin, China.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 14 April 2021
doi: 10.3389/fpsyg.2021.657660

Check for updates

Further Support for the Psychometric Properties of the Farsi Version of Perth Alexithymia Questionnaire

Arezou Lashkari[1], Mohsen Dehghani[1], Vahid Sadeghi-Firoozabadi[1], Mahmood Heidari[1] and Ali Khatibi[2,3]**

[1] Department of Psychology, Shahid Beheshti University, Tehran, Iran, [2] Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom, [3] Centre for Human Brain Health, University of Birmingham, Birmingham, United Kingdom

Alexithymia is defined as the lack of words to describe emotions and is associated with different psychopathologies. Various tools have been developed for measuring alexithymia; each has its limitations. A new questionnaire, Perth Alexithymia Questionnaire (PAQ), was developed to simultaneously assess positive and negative dimensions. Validation of such a tool in different cultures allows cross-cultural health psychology studies and facilitates knowledge transfer in the field. We aimed to examine the psychometric features of the PAQ in the Farsi-speaking population in Iran. Four-hundred-twenty-nine university students were asked to complete the PAQ, the Toronto Alexithymia Scale (TAS-20), Beck Depression Inventory (BDI-II), Beck Anxiety Inventory (BAI), and emotion regulation questionnaire (ERQ). Concurrent validity, discriminant validity, internal consistency, and test-retest reliability and factor structure were investigated. Confirmatory factor analysis showed a five-factor model identical to the original questionnaire. The questionnaire indicated good internal consistency ($0.82 < \alpha < 0.94$). Test-retest reliability was acceptable for all subscales. The correlations between PAQ and its subscales with BDI-II, BAI, and TAS, and expression suppression subscale of ERQ were strong for concurrent validity. Concerning the discriminant validity, PAQ and its subscales were not correlated with reappraisal subscales of ERQ. The present findings suggest that the Farsi version of PAQ has strong psychometric properties and is appropriate for use in the Farsi-speaking population.

Keywords: alexithymia, psychometric properties, validity, reliabiity, Farsi (Persian)

Edited by:
Paola Gremigni,
University of Bologna, Italy

Reviewed by:
Sina Hafizi,
University of Manitoba, Canada
Rory Allen,
Goldsmiths University of London, United Kingdom

****Correspondence:***
Mohsen Dehghani
m.dehghani@sbu.ac.ir
Ali Khatibi
Ali.khatibi@gmail.com

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *23 January 2021*
Accepted: *23 March 2021*
Published: *14 April 2021*

Citation:
Lashkari A, Dehghani M, Sadeghi-Firoozabadi V, Heidari M and Khatibi A (2021) Further Support for the Psychometric Properties of the Farsi Version of Perth Alexithymia Questionnaire. Front. Psychol. 12:657660. doi: 10.3389/fpsyg.2021.657660

INTRODUCTION

Sifneos (1973) introduced alexithymia by describing psychosomatic patients who could not find appropriate words to express their emotions. Psychosomatic disorders have been described in various forms in previous versions of the Diagnostic and Statistical Manual of Mental Disorders (DSM). DSM-I described psychosomatic disorders in a section called psychophysiological autonomic and visceral disorders. DSM 5 refers to psychosomatic disorders in Somatic Symptom and Related Disorders (Moldovan et al., 2015). In all disorders of somatic symptom and related disorders, physical symptoms along with distress and impairment are prominent (American Psychiatric and Association [APA]., 2013). Alexithymia is one of the salient features of psychosomatic disorders. These patients show somatic symptoms instead of expressing emotions. Although alexithymia was coined first as a standard feature in psychosomatic disorders, it is prevalent in the normal population (Honkalampi et al., 2000; Kokkonen et al., 2001). Most researchers proposed that alexithymia is related to psychological pathologies. For example, Honkalampi et al. (2001) stated that the severity of depression was associated with alexithymia. de Bruin et al. (2019) showed that substance use disorder (SUD) patients with post-traumatic stress disorder (PTSD) are more alexithymic than SUD patients without PTSD. Iskric et al. (2020) suggest that alexithymia is related to non-suicidal self-injury and suicidal ideation. Passardi et al. (2019) indicated that alexithymia played an essential role in facial emotion recognition of negative emotions in PTSD. In addition, McCallum et al. (2003) showed that alexithymia could affect the treatment outcomes by interfering with the improvement of general symptoms.

These studies demonstrated the importance of alexithymia as a construct. Toronto Alexithymia Scale (TAS) (Bagby et al., 1994) and Bermond-Vorst Alexithymia Questionnaire (BVAQ) (Vorst and Bermond, 2001) are the two most commonly used measures. TAS includes difficulty describing feelings (DDF), difficulty identifying feelings (DIF), and externally oriented thinking (EOT). TAS-20 is one of the first developed measures of alexithymia for both clinical and non-clinical samples. However, some researchers claimed that the EOT subscale of TAS does not have enough internal consistency (Kojima et al., 2001; Taylor et al., 2003; Bagby et al., 2020), and items load poorly on the related latent factor (Preece et al., 2018a,b). TAS correlated positively with negative affect measures and raised concern that TAS might assess distress rather than alexithymia construct (Preece et al., 2020). The other criticism of TAS is that it does not take the valence of emotions into account (Preece et al., 2018b).

Bermond-Vorst Alexithymia Questionnaire, another commonly used tool, assesses cognitive and affective dimensions of alexithymia. The cognitive dimension of alexithymia refers to the processing of emotions at a cognitive level. The affective dimension measures the subjective report of individual emotional experience. Emotionalizing and fantasizing characterize affective alexithymia. Alexithymia's cognitive dimension involve analyzing, identifying, and verbalizing emotions (Bermond et al., 2007). Differentiating the two dimensions of alexithymia is considered controversial (Goerlich, 2018). Some studies have failed to show the ability to differentiate between the two dimensions (Watters et al., 2016). Some researchers argued that difficulty in emotionalizing is not a valid indicator of alexithymia, as it does not distinguish between negative and positive reactivity to emotions (Preece et al., 2017).

Perth Alexithymia Questionnaire (PAQ; Preece et al., 2018b) was developed to address the earlier tool's limitations. PAQ is a 24-item questionnaire containing 5-subscales: positive DIF (P-DIF), negative DIF (N-DIF), positive DDF (P-DDF), negative DDF (N-DDF), and general EOT (G-EOT). PAQ is developed based on the attention-appraisal model of emotions (Gross, 2015). The model contains four stages: situation, attention, evaluation, and response. When an emotional response becomes the stimulus (situation stage) that is the target of valuation, the person pays attention to it (attention stage) and starts to evaluate it (appraisal). Then, the individual might respond to it (response stage). According to the model, alexithymic people cannot focus their attention on emotional response. Furthermore, individuals with high levels of alexithymia cannot evaluate the emotional response as what it is and what it means. As a result, DIF and DDF are indicators of the appraisal stage. The salient point is that the appraisal stage can consist of valence. Both DIF and DDF subscales of PAQ comprise positive and negative valences and used to calculate the difficulty in the appraisal feeling (DAF) composite subscale. By combining the 5-subscale, composite subscales can be produced. In this regard, positive difficulty appraising feeling (P-DAF) includes positive DIF and positive DDF. Negative difficulty appraising feeling (N-DAF) is the sum of negative DIF and negative DDF (Preece et al., 2018b).

As mentioned, PAQ is constructed based on the attention-appraisal model of emotions. Most studies revealed that alexithymia reflects a deficit in regulating emotions (Swart et al., 2009; Pandey et al., 2011). Also, it has been suggested that alexithymic individuals used more suppression and less reappraisal for emotion regulation (Swart et al., 2009). Cognitive reappraisal means cognitive changes in the affective influence of the emotion-eliciting situation. Expressive suppression inhibits emotional behavior. Cognitive reappraisal is assumed as an adaptive emotion regulation strategy. However, expressive suppression is indicative of a maladaptive emotion regulation strategy (Gross and John, 2003). It is shown that emotional expression is strongly related to the cultural context (Altarriba and Kazanas, 2017). So, it is crucial to assess the psychometric features of the measures in other cultures.

Research showed the role of culture in factorial structure (Fernández-Jiménez et al., 2013). TAS-20 and more recently PAQ are measures for assessing the alexithymia in Iran. As we know, EOT component of the Farsi version of TAS-20 has lower internal consistency (Besharat, 2008). Besides, other studies aimed at validating PAQ in Iran failed to recruit a big enough sample or to report confirmatory analysis (Heydari et al., 2020; Mousavi Asl et al., 2020). Accordingly, the current study aimed to examine the validity and reliability of PAQ in Iran with specific reference to its internal consistency, test-retest reliability, concurrent and discriminant validity, and factor structure. In this regard, depression, anxiety inventories and TAS, suppression subscale of emotion regulation questionnaire (ERQ) were used for assessing concurrent validity. Discriminant validity was assessed by the cognitive reappraisal subscale of the ERQ. Confirmatory factor analyses of PAQ tested the construct validity. Considering the psychometric properties of the original version of the PAQ, we expected the measure to be highly reliable. We hypothesized that PAQ is an excellent measure to assess alexithymia as reflected in its correlation with other measures of emotional and mood difficulties. We also expected the PAQ to have a unique contribution above that of the previous measures as reflected in factor analysis.

Method

Participants

Participants were recruited via advertisements on notice boards in the universities. They were included if they reported a good level of literacy, Including reading and understanding Farsi.

Those with a history of traumatic experiences in the previous 6 months to the data collection point were excluded. A total of 436 individuals answered the questionnaires. Incomplete questionnaires were removed from the final sample leading to 429 (63.2% female, mean age = 21.35 ± 2.88, range = 18–39) unique responses that were included in the statistical analyses. All participants were university students, including 364 undergraduate, 41 masters, 11 Ph.D. students. The study was approved by the Research Ethics Committee of Shahid Beheshti University, Tehran, Iran.

MATERIALS AND METHODS

Perth Alexithymia Questionnaire

Perth Alexithymia Questionnaire (Preece et al., 2018b) consists of 24 self-report measures rated on a seven-point Likert scale (1 = strongly disagree; 7 = strongly agree), with high scores demonstrating high levels of alexithymia. PAQ, the only alexithymia measure including the valence of emotions, consists of five subscales: negative-difficulty identifying feelings (N-DIF), positive-difficulty identifying feelings (P-DIF), negative-difficulty describing feelings (N-DDF), positive-difficulty describing feelings (P-DDF), and general-externally orientated thinking (G-EOT). "When I am feeling bad, I can't tell whether I'm sad, angry or scared" and "when I'm feeling good, I get confused about what emotion it is" are the samples of N-DIF and P-DIF. The subscales can combine and produce the composite subscales. PAQ demonstrated good concurrent and discriminant validity and good internal consistency (Cronbach's alpha of the subscales ranged from 0.87 to 0.96) (Preece et al., 2018b).

We followed a standard translation and back-translation procedure for the introduction of the Farsi version of the PAQ. First, the original questionnaire was translated into Farsi, and then a bilingual professional psychologist back-translated it into English. The back-translated version was compared with the original English one, a few minor corrections were applied and the final version of the PAQ, used in this study, was reached.

20-Item Toronto Alexithymia Scale

20-Item Toronto Alexithymia Scale (Bagby et al., 1994) is a 20-item measure consisting of three subscales: difficult identifying feelings (DIF), DDF, and EOT style. Each item is rated on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). "When I am upset, I don't know if I am sad, frightened or angry", "It is difficult for me to find the right words for my feelings", and "Being in touch with emotions is essential" are the samples of DIF, DDF, and EOT, respectively. Five key items are reverse scored (4, 5, 10, 118, and 19). TAS-20 demonstrated good internal consistency (Cronbach's alpha ranged from 0.66 to 0.75) and test-retest reliability (0.77). The three-factor structure was congruent with the theoretical model underlying TAS-20 (Bagby et al., 1994). TAS psychometric properties have been tested and confirmed by various studies (Bressi et al., 1996; Taylor et al., 2003; Säkkinen et al., 2007). The Farsi version of TAS demonstrated good validity and reliability in the Farsi-speaking population (Besharat, 2008).

Emotion Regulation Questionnaire

Emotion regulation questionnaire (Gross and John, 2003) is a short questionnaire designed to separate two subscales: expressive suppression (I keep my emotions to myself) and cognitive reappraisal (When I want to feel more positive emotion [such as joy or amusement], I change what I'm thinking about). ERQ is answered on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The Cronbach's alpha was found to be in the range of 0.68–0.82 in different populations. ERQ demonstrated good convergent and discriminant validity (Gross and John, 2003). ERQ showed good validity and reliability in the Farsi-speaking population. The Cronbach's alpha coefficient for cognitive reappraisal and expressive suppression were 0.78 and 0.60, respectively (Ghasempour et al., 2012).

Beck Depression Inventory-II

Beck depression inventory-II (Beck et al., 1996) is a widely used measure for assessing depression in both clinical and non-clinical populations. BDI-II contains 21 items, each answer being scored on a Likert scale value of 0–3 in which higher scores indicate the existence of depression symptoms. One of the scale items is as follows: I do not feel sad, I feel sad much of the time, I am sad all the time, I am so sad or unhappy that I can't stand it. Individuals are asked to select one of four possible items in each question based on the last 2 weeks state. The internal consistency is reported around 0.90, and test-retest reliability from 0.73 to 0.96. BDI-II demonstrated two-factor of cognitive-affective and somatic-vegetative (Wang and Gorenstein, 2013). In the Farsi-speaking sample, the same two-factor model is confirmed (Ghassemzadeh et al., 2005), and internal consistency and test-retest reliability were 0.87 and 0.74.

Beck Anxiety Inventory

The BAI (Beck et al., 1988) is a 21-item self-report questionnaire that asks about common anxiety symptoms such as being scared, nervous, and unsteady. The items are rated on a 4-point Likert scale (0 = not at all, 3 = severely). The total score ranged from 0 to 63. The higher scores indicate a higher level of anxiety. BAI has the clinical classification, with scores from 0 to 7 displaying minimal anxiety, 8–15 as mild anxiety, 16–25 as moderate anxiety, and 26–63 as severe anxiety. Scores greater than 16 indicate clinically significant anxiety (Beck et al., 1988). BAI internal consistency (alpha Cronbach) and test-retest reliability reported as 0.91 and 0.65, respectively. BAI demonstrated a two-factor solution and concurrent validity with other anxiety measures and the BDI-II as well (Bardhoshi et al., 2016). In the Iranian population, the internal consistency coefficient (alpha Cronbach) was around 0.90, and the three-factor model is confirmed (Dobson and Mohamadkhani, 2008).

Data Analysis

Confirmatory factor analyses (CFA) was performed using LISREL 8.80. All other analyses, such as correlations and descriptive statistics, were done by SPSS 24. As mentioned before, the PAQ concept is based on the attention-appraisal model of alexithymia. As a result, the subscale of PAQ can combine and produce composite subscales according to the theory (Preece et al., 2018b). Five theoretical models for the factor structure have been suggested. In this study, CFA was applied for all five models. CFA was conducted using the maximum likelihood estimation based on the Pearson covariance matrix. Maximum likelihood requires that the data display both univariate and multivariate normality, despite the robustness to violation of normality as long as the sample size is large (Gerbing and Anderson, 1985; Finney and DiStefano, 2006).

The fitness of the model was evaluated with the most important indices such as root mean square error of approximation (RMSEA), comparative fit index (CFI), normed fit index (NFI), incremental fit index (IFI), standard root mean squared residual (SRMR), Akaike information criterion (AIC), and relative chi-square (χ^2/df). Acceptable fit values of CFI, NFI, and IFI are greater than 0.90, while the acceptance value of χ^2/df is less than 5 (Hu and Bentler, 1999; Hooper et al., 2008). The acceptable value of RMSEA and SRMR is less than 0.08 and 0.1, respectively (Schumacker and Lomax, 2016). AIC is a criterion to compare different models, and lower values demonstrate better fitness (Kline, 2015).

Cronbach's alpha was calculated to assess internal consistency. The values greater than 0.90 indicate excellent consistency, while those greater than 0.80 and 0.70 indicate good and acceptable internal consistency, respectively (Groth-Marnat, 2009). The test-retest reliability was examined in a separate group of 59 participants (76.3% female) with a mean age of 27.49 (SD = 5.12, range = 18–47) that were recruited only for this purpose. They completed the PAQ questionnaire two times with a 2-week interval between them. The test-retest reliability was quantified using a Pearson correlation estimate for total scores.

The Pearson correlation coefficient was used to assess concurrent and discriminant validity. We hypothesized that a medium to large (around 0.3 and 0.5) correlation (Cohen, 1998) between PAQ and TAS, BDI-II, BAI, and expressive suppression subscale of ERQ indicates concurrent validity. We expected a small, around 0.1 (Cohen, 1998) and even negative correlation of PAQ and cognitive reappraisal subscale of ERQ as discriminant validity.

Procedure

At the beginning of the session, participants received the consent form. After reading and accepting to participate, they received the battery of questionnaires and were instructed to complete them by paying attention to the instruction on top of each questionnaire. The researcher (AL) was accessible during the time they were completing the questionnaires. At the end of the session, participants were debriefed. For the test-retest, due to the restriction imposed as the consequence of the COVID19 pandemic, a separate group of participants received a link with the online format of the translated PAQ questionnaire. They received a similar link 2 weeks later with the same questionnaire and were asked to answer it within 24 h.

RESULTS

Descriptive Statistics

Table 1 presents the mean and standard deviation for subscales and total PAQ, and other measures in the study. The minimum and maximum mean and standard deviation belonged to the P-DIF and the G-EOT subscales, respectively (M = 9.84, SD = 4.96; M = 21.38, and SD = 9.57).

Confirmatory Factor Analyses

Before applying the CFA, all 24 items were checked for normal distribution. A skewness value of ± 1 and ± 2 is considered excellent and acceptable, respectively. While a value of ± 3 is described as highly skewed. Kurtosis greater than 10 indicates non-normal distribution (George and Mallery, 2006; Kline, 2015). In the present study, the skewness ranged between 0.24 and 1.52, and the range of kurtosis was from -1.15 to 1.47, indicating a normal distribution of the scores.

Model 1 was a one-factor model in which all 24 items were loaded in general alexithymia. Model 2 was a two-factor model in which items were loaded on G-EOT and G-DAF. This model discriminates attention and appraisal stages of emotion evaluation. Model 3 was a three-factor model consisting of

TABLE 1 | Descriptive statistics for the administered measures.

Measure/subscale		Total	
		M	SD
Perth Alexithymia Questionnaire	P-DIF	9.84	4.96
	N-DIF	12.40	5.91
	P-DDF	11.05	5.83
	N-DDF	13.32	6.49
	G-EOT	21.38	9.57
	G-DIF	22.25	9.96
	G-DDF	24.37	11.27
	N-DAF	25.72	11.72
	P-DAF	20.89	10.27
	G-DAF	46.62	20.28
	ALEXI	68.01	27.05
TAS	DIF	15.71	5.69
	DDF	13.27	4.34
	EOT	18.94	3.90
ERQ	Expressive and suppression	13.23	5.42
	Cognitive and reappraisal	26.34	7.33
BAI		11.86	9.49
BDI-II		14.62	10.488

M, mean; SD, standard deviation; PAQ, Perth Alexithymia Questionnaire; N-DIF, negative-difficulty identifying feelings; P-DIF, positive-difficulty identifying feelings; NDDF, negative-difficulty describing feelings; P-DDF, positive-difficulty describing feelings; G-EOT, general-externally orientated thinking; GDIF, general-difficulty identifying feelings; G-DDF, general-difficulty describing feelings; N-DAF, negative-difficulty appraising feelings; PDAF, positive-difficulty appraising feelings; DAF, general-difficulty appraising feelings; ALEXI, alexithymia; BAI, beck anxiety inventory; BDI, beck depression inventory; TAS, Toronto alexithymia questionnaire; DIF, difficulty identifying the feeling; DDF, difficulty describing the feeling; EOT, externally oriented thinking; and ERQ, emotion regulation questionnaire.

G-DIF, G-DDF, and G-EOT. G-DIF is obtained by combining positive and negative DIF, and the G-DDF is obtained by summing the positive DDF and negative DDF. There was no distinction between the valence of components in this model. Model 4 is a three-factor model based on the distinction of valence and includes G-EOT, N-DAF, and P-DAF factors. N-DAF is referred to as negative DIF, and negative DDF and P-DAF are created by collecting positive DIF and positive DDF. Model 5 is a five-factor model, and all factors were divided based on valence, DIF, and DDF. The items were loaded on G-EOT, N-DIF, P-DIF, N-DDF, and P-DDF. The goodness of fit values in **Table 2** indicates that models 5 and 1 revealed the best and poorest fit indices, respectively. All the values of model 5 demonstrate that model 5 is the best-fitted model comparing other models. All items were loaded well on five hypothesized factors (**Figure 1**).

Reliability

We examined the reliability of the measure in two different ways. First, we measured the test-retest reliability for the subscales and total scales, which are presented in **Table 3** (left column). Test-retest reliability of all subscales was statistically significant after at least 2 weeks ($r > 0.69$). Cronbach's alpha coefficient ranged from 0.82 to 0.94 for subscales and 0.94 for the total score, suggesting the good internal consistency of PAQ.

Concurrent and Discriminant Validity

Table 4 presents the correlation of all subscales of PAQ and TAS, ERQ, BAI, and BDI-II. As expected, the correlations between all subscales of PAQ and TAS are statistically significant, ranging from 0.18 to 0.73. Also, there are positive correlations between PAQ and BDI-II and BAI. The expressive suppression subscale of ERQ and PAQ subscales were significantly correlated (0.37–0.53), and no significant correlation was found between PAQ and cognitive reappraisal.

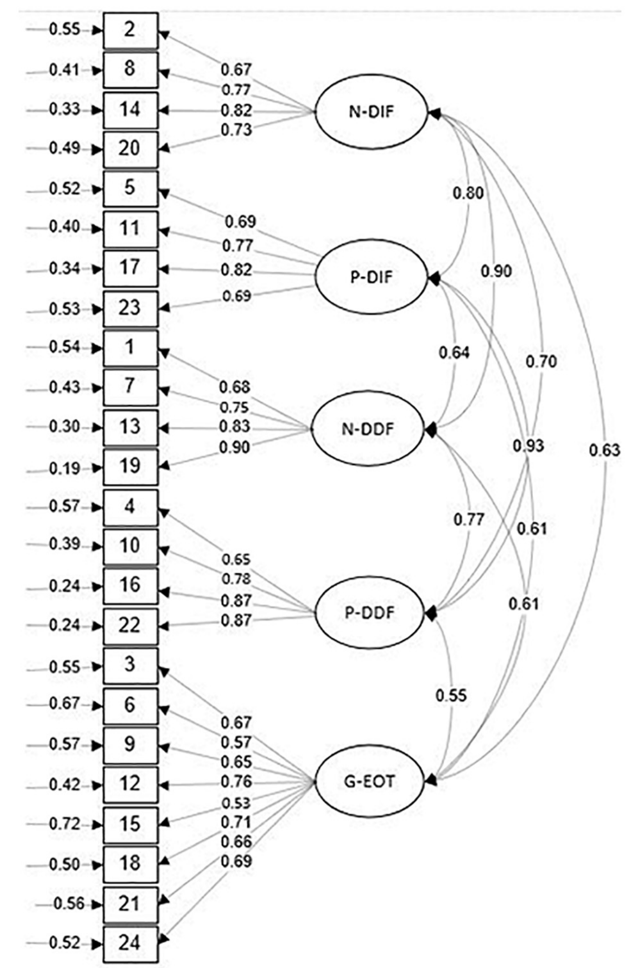

FIGURE 1 | Confirmatory factor analysis: Item loadings on the 5-factor model, negative-difficulty identifying feelings (N-DIF), positive-difficulty identifying feelings (P-DIF), negative-difficulty describing feelings (N-DDF), positive-difficulty describing feelings (P-DDF), and general-externally orientated thinking (G-EOT).

DISCUSSION

This study aimed to investigate the utility of PAQ in Farsi-speaking people living in Iran. We aimed to determine which of the original model's five variants would fit best in the Farsi-speaking population. Although most of the fit indices were acceptable in all five models, model 5 (which consists of five-factors) was the best. In model 5, all 24 items were loaded on the related factors. Therefore, it can be claimed that the Farsi version of the PAQ is a multidimensional questionnaire. This claim is consistent with the results of Preece et al. (2018b). As assumed by prior studies (Barrett et al., 2001), differences between positive and negative emotions are important for emotion regulation because the individuals differ in their emotional experiences. One of the most distinctive features of the PAQ is that it considers the valence of emotions, which is well described in separating the positive and negative valence in apprising factors. P-DIF, N-DIF, P-DDF, and N-DDF are the indicators of the appraisal stage of the model. Model 5 contains all four mentioned factors plus the EOT, which is not valence specified.

Testing the measure's reliability, we found the PAQ subscale and the composite subscales to be reliable constructs similar

TABLE 2 | Goodness-of-fit index values from confirmatory factor analysis of the 24 Perth Alexithymia Questionnaire items.

Model	Chi-Square	Df	$\chi 2/df$	RMSEA	CFI	NFI	IFI	AIC	SRMR
Model1	2,857.29	252	11.33	0.15	0.91	0.90	0.91	2,953.29	0.09
Model2	1,892.36	251	7.53	0.12	0.94	0.93	0.94	1,990.36	0.06
Model3	1,928.67	249	7.74	0.12	0.94	0.93	0.94	2,030.15	0.06
Model4	10.56	249	4.24	0.08	0.97	0.95	0.97	1,158.67	0.04
Model5	784.53	242	3.24	0.07	0.97	0.96	0.97	900.53	0.04

Df, the degree of freedom; χ^2/df, relative chi-square; RMSEA, root mean square error; CFI, the comparative fit index; NFI, normed fit index; IFI, incremental fit index; SRMR, standard root mean squared residual; and AIC, Akaike information criterion.

to the original version (Preece et al., 2018b). The purpose of creating the new measure was to overcome the limitations of previous tools, including the low reliability of the EOT subscale.

TABLE 3 | Test-retest reliability and Cronbach's alpha reliability coefficients for administered measures.

Measure/subscales		Test-retest correlation	Cronbach's alpha (α)
PAQ	P-DIF	0.71**	0.82
	N-DIF	0.72**	0.83
	P-DDF	0.69**	0.87
	N-DDF	0.71**	0.87
	G-EOT	0.74**	0.85
	G-DIF	0.76**	0.88
	G-DDF	0.77**	0.90
	N-DAF	0.75**	0.91
	P-DAF	0.76**	0.91
	G-DAF	0.81**	0.94
	ALEXI	0.85**	0.94

PAQ, Perth Alexithymia Questionnaire; N-DIF, negative-difficulty identifying feelings; P-DIF, positive-difficulty identifying feelings; NDDF, negative-difficulty describing feelings; P-DDF, positive-difficulty describing feelings; G-EOT, general-externally orientated thinking; GDIF, general-difficulty identifying feelings; G-DDF, general-difficulty describing feelings; N-DAF, negative-difficulty appraising feelings; PDAF, positive-difficulty appraising feelings; DAF, general-difficulty appraising feelings; ALEXI, alexithymia.
$^{**}p < 0.01$.

TABLE 4 | Pearson correlations between the Perth Alexithymia Questionnaire, beck depression inventory, beck anxiety inventory, emotion regulation questionnaire, and Toronto Alexithymia Scale.

	Measures/ subscales	BDI-II	BAI	ERQ		TAS-		
				Cognitive reappraisal	Expressive suppression	DIF	DDF	EOT
PAQ	Subscales							
	N-DIF	0.39**	0.33**	0.03	0.39**	0.73**	0.57**	0.18**
	P-DIF	0.28**	0.21**	−0.04	0.37**	0.62**	0.50**	0.28**
	N-DDF	0.35**	0.20**	0.01	0.44**	0.58**	0.67**	0.19**
	P-DDF	0.26**	0.16**	−0.07	0.41**	0.52**	0.58**	0.22**
	G-EOT	0.34**	0.21**	0.02	0.53**	0.47**	0.49**	0.28**
	Composites							
	G-DIF	0.37**	0.30**	−0.01	0.41**	0.74**	0.59**	0.25**
	G-DDF	0.34**	0.20**	0.00	0.46**	0.61**	0.69**	0.22*
	N-DAF	0.39**	0.28**	0.02	0.44**	0.69**	0.66**	0.20**
	P-DAF	0.28**	0.19**	−0.02	0.41**	0.60**	0.58*	0.26**
	G-DAF	0.37**	0.26**	−0.04	0.46**	0.70**	0.67**	0.25**
	ALEXI	0.40**	0.27**	−0.01	0.53**	0.70**	0.68**	0.28**

PAQ, Perth Alexithymia Questionnaire; N-DIF, negative-difficulty identifying feelings; P-DIF, positive-difficulty identifying feelings; NDDF, negative-difficulty describing feelings; P-DDF, positive-difficulty describing feelings; G-EOT, general-externally orientated thinking; GDIF, general-difficulty identifying feelings; G-DDF, general-difficulty describing feelings; N-DAF, negative-difficulty appraising feelings; PDAF, positive-difficulty appraising feelings; DAF, general-difficulty appraising feelings; ALEXI, alexithymia; BAI, beck anxiety inventory; BDI, beck depression inventory; TAS, Toronto alexithymia questionnaire; DIF, difficulty identifying the feeling; DDF, difficulty describing the feeling; EOT, externally oriented thinking; and ERQ, emotion regulation questionnaire.
$^{**}(p < 0.01)$; $^{*}(p < 0.05)$.

Most studies presented that EOT component has low reliability in different cultures (Bressi et al., 1996; Kojima et al., 2001; Taylor et al., 2003). In the present study, all the alpha coefficients were greater than 0.80, reflecting good internal consistency, particularly EOT subscale, which is consistent with the study by Preece et al. (2018b). The original study did not assess the test-retest reliability (Preece et al., 2018b). All subscales, especially the total alexithymia, indicated good test-retest reliability, which means that the scale results are consistent at different time points. The test-retest reliability results in the present study are comparable with those reported by Mousavi Asl et al. (2020).

To test the concurrent and discriminant validity of the PAQ, we measured the correlations between the subscales and other previously validated measures, including the BDI-II, BAI, TAS, and ERQ. In agreement with the original study (Preece et al., 2018b), PAQ was strongly correlated with all TAS subscales, which is another alexithymia measure. A score for general difficulties in identifying and describing feelings will be obtained by combining the negative and positive subscales. These two subscales, along with the externally orientated thinking subscale, are the same subscales that TAS measures too. Therefore, a high correlation between the two measures was expected. The correlation of PAQ with BDI-II and BAI is in line with the original study (Preece et al., 2018b). Individuals with depression and anxiety seemed to have difficulty recognizing and describing their emotions. Honkalampi et al. (2001) stated that alexithymia is associated with depression. Moreover, Honkalampi et al. (2000) reported that the prevalence of alexithymia in individuals with mild depression (BDI > 9) was 32%. Another study showed that difficulties in describing and identifying feelings, changes with mood and recovery from depression and were associated with a decrease in alexithymia (Saarijärvi et al., 2001). Alexithymia is also associated with anxiety. Difficulties describing and identifying feelings are related to anxiety disorders such as generalized anxiety disorder (GAD), and the presence of alexithymia is related to higher levels of anxiety (Berardis et al., 2008).

The expressive suppression subscale of the ERQ was correlated positively with the PAQ subscales. However, PAQ was not correlated with the cognitive reappraisal subscale of ERQ. Cognitive reappraisal and expressive suppression are considered adaptive and maladaptive emotion regulation, respectively (Gross and John, 2003). As mentioned before, alexithymia is defined as a deficiency in expressing emotions. Accordingly, it is expected that alexithymia would be correlated with expressive suppression. In line with the lack of relationship between alexithymia and cognitive reappraisal, it can be stated that alexithymia is an obstacle to regulating emotions (Gross, 1998). It can be assumed that the alexithymia has concurrent and discriminant validity.

The present study, despite its advantages, also suffers from some limitations. Firstly, the study was carried out among university students. Therefore, generalizing the finding to other populations, especially clinical samples, should be done with caution. Secondly, the cutoff point was not assessed, and this will limit the use of the measure for clinical research. Hence, for future studies, it is recommended to evaluate the psychometric feature of the PAQ in other populations, especially clinical samples.

Understanding emotion and its contribution to wellbeing is important in health psychology research. Because of the universality of emotions, they can help us in cross-cultural studies. Adaptation of the existing tools to new languages and in different cultures can facilitate cross-cultural health psychology research (Rudell and Diefenbach, 2008). PAQ (Preece et al., 2018b) is one of the most widely used tools to test alexithymia. Previous efforts in adapting and validating this tool among the Farsi-speaking population suffer from low sample size and lack of model estimation. In the current study, we examined the psychometric properties of a new adaptation of the PAQ (please see **Supplementary Material**) in a large Farsi-speaking sample in Iran. In summary, the study findings show that the Farsi version of PAQ is a valid and reliable measure. PAQ seems to be a promising measure for identifying the deficiency of emotions in the Farsi-speaking sample.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Department of Psychology, Shahid Beheshti University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AL was involved in design, data collection, analysis, and writing. MD and AK was involved in design and writing. VS-F and MH was involved in design. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.657660/full#supplementary-material

REFERENCES

Altarriba, J., and Kazanas, S. A. (2017). "Emotions and Expressions across Cultures," in *The International Encyclopedia of Intercultural Communication*, eds Y. Y. Kim and K. McKay-Semmler (Hoboken, NJ: John Wiley & Sons, Inc), 1–10. doi: 10.1002/9781118783665.ieicc0247

American Psychiatric and Association [APA]. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5§)*. Washington, DC: American Psychiatric Association.

Bagby, R. M., Parker, J. D. A., and Taylor, G. J. (2020). Twenty-five years with the 20-item Toronto Alexithymia Scale. *J Psychosom Res* 131, 109940. doi: 10.1016/j.jpsychores.2020.109940

Bagby, R. M., Parker, J. D., and Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *J Psychosom Res* 38, 23–32. doi: 10.1016/0022-3999(94)90005-1

Bardhoshi, G., Duncan, K., and Erford, B. T. (2016). Psychometric meta-analysis of the English version of the Beck Anxiety Inventory. *J Couns Dev* 94, 356–373. doi: 10.1002/jcad.12090

Barrett, L. F., Gross, J., Christensen, T. C., and Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cogn Emot* 15, 713–724. doi: 10.1080/02699930143000239

Beck, A. T., Epstein, N., Brown, G., and Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol* 56, 893. doi: 10.1037/0022-006x.56.6.893

Beck, A. T., Steer, R. A., and Brown, G. K. (1996). *Manual for the beck depression inventory-II*. San Antonio, TX: Psychological Corporation.

Berardis, D., Campanella, D., Nicola, S., Gianna, S., Alessandro, C., Chiara, C., et al. (2008). The Impact of Alexithymia on Anxiety Disorders: a Review of the Literature. *Curr Psychiatry Rev* 4, 80–86. doi: 10.2174/157340008784529287

Bermond, B., Clayton, K., Liberova, A., Luminet, O., Maruszewski, T., Ricci Bitti, P. E., et al. (2007). A cognitive and an affective dimension of alexithymia in six languages and seven populations. *Cogn Emot* 21, 1125–1136. doi: 10.1080/02699930601056989

Besharat, M. A. (2008). Psychometric characteristics of Persian version of the Toronto alexithymia scale-20 in clinical and non-clinical samples. *Iran J Med Sci* 33, 1–6.

Bressi, C., Taylor, G., Parker, J., Bressi, S., Brambilla, V., Aguglia, E., et al. (1996). Cross validation of the factor structure of the 20-item Toronto Alexithymia Scale: an Italian multicenter study. *J Psychosom Res* 41, 551–559. doi: 10.1016/s0022-3999(96)00228-0

Cohen, J. (1998). *Statistical power analysis for the behavioral sciences*. New Jersey, NJ: Lawrence erlbaum association.

de Bruin, P. M. J., de Haan, H. A., and Kok, T. (2019). The prediction of alexithymia as a state or trait characteristic in patients with substance use disorders and PTSD. *Psychiatry Res* 282, 1–6. doi: 10.1016/j.psychres.2019.112634

Dobson, K., and Mohamadkhani, P. (2008). Psychometric properties of Beck Depression Inventory- II in a sample of people with depressive disorder. *IRJ* 39, 80–89.

Fernández-Jiménez, E., Pérez-San-Gregorio, M. Á, Taylor, G. J., Michael Bagby, R., Ayearst, L. E., and Izquierdo, G. (2013). Psychometric properties of a revised Spanish 20-item Toronto Alexithymia Scale adaptation in multiple sclerosis patients. *Int J Clin Health Psychol* 13, 226–234. doi: 10.1016/s1697-2600(13)70027-9

Finney, S. J., and DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Hrsg.). *Structural equation modeling: a second course*. Greenwich, CT: Information Age Publishing, 269–314.

George, D., and Mallery, P. (2006). *IBM SPSS statistics 26 step by step: a simple guide and reference*. New York, NY: Taylor & Francis.

Gerbing, D. W., and Anderson, J. C. (1985). The Effects of Sampling Error and Model Characteristics on Parameter Estimation for Maximum Likelihood Confirmatory Factor Analysis. *Multivariate Behav Res* 20, 255–271. doi: 10.1207/s15327906mbr2003_2

Ghasempour, A., Eilbeigi, R., and Hassanzadeh, S. (2012). "Psychometric properties of the emotional regulation questionnaire of Gross and John in an Iranian sample," in *The Sixth Congress on Mental Health of University Students*. Gilan-Iran.

Ghassemzadeh, H., Mojtabai, R., Karamghadiri, N., and Ebrahimkhani, N. (2005). Psychometric properties of a Persian-language version of the Beck Depression Inventory–Second edition: BDI-II-PERSIAN. *Depress Anxiety* 21, 185–192. doi: 10.1002/da.20070

Goerlich, K. S. (2018). The Multifaceted Nature of Alexithymia - A Neuroscientific Perspective. *Front. Psychol* 9:1614. doi: 10.3389/fpsyg.2018.01614

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Rev Gen Psychol* 2, 271–299. doi: 10.1037/1089-2680.2.3.271

Gross, J. J. (2015). Emotion Regulation: Current Status and Future Prospects. *Psychol Inq* 26, 1–26. doi: 10.1080/1047840x.2014.940781

Gross, J. J., and John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *J Pers Soc Psychol* 85, 348–362. doi: 10.1037/0022-3514.85.2.348

Groth-Marnat, G. (2009). *Handbook of psychological assessment*. New Jersey, NJ: John Wiley & Sons.

Heydari, S., Lajmiri, S., Azadyekta, M., Barzegar, M., and Arshadi, M. (2020). Reliability and validity of Perth alexithymia questionnaire (PAQ) and its relation with cognitive emotion regulation. *Iranian Journal of Rooyesh_e_Ravanshenasi* 9, 73–80.

Honkalampi, K., Hintikka, J., Laukkanen, E., Lehtonen, J., and Viinamaki, H. (2001). Alexithymia and depression: a prospective study of patients with major depressive disorder. *Psychosomatics* 42, 229–234. doi: 10.1176/appi.psy.42.3.229

Honkalampi, K., Hintikka, J., Tanskanen, A., Lehtonen, J., and Viinamäki, H. (2000). Depression is strongly associated with alexithymia in the general population. *J Psychosom Res* 48, 99–104. doi: 10.1016/s0022-3999(99)00083-5

Hooper, D., Coughlan, J., and Mullen, M. (2008). Structural equation modelling: guidelines for determining model fit. *Electron J Bus Res Methods Vol* 6, 53–60.

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 6, 1–55. doi: 10.1080/10705519909540118

Iskric, A., Ceniti, A. K., Bergmans, Y., McInerney, S., and Rizvi, S. J. (2020). Alexithymia and self-harm: A review of nonsuicidal self-injury, suicidal ideation, and suicide attempts. *Psychiatry Res* 288, 1–15. doi: 10.1016/j.psychres.2020.112920

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York, NY: Guilford publications.

Kojima, M., Frasure-Smith, N., and Lespérance, F. (2001). Alexithymia following myocardial infarction: psychometric properties and correlates of the Toronto Alexithymia Scale. *J Psychosom Res* 51, 487–495.

Kokkonen, P., Karvonen, J. T., Veijola, J., Läksy, K., and Jokelainen, J. (2001). Perceived and sociodemographic correlates of alexithymia in a population sample of young adults. *Compr Psychiatry*. 42, 471–476. doi: 10.1053/comp.2001.27892

McCallum, M., Piper, W. E., Ogrodniczuk, J. S., and Joyce, A. S. (2003). Relationships among psychological mindedness, alexithymia and outcome in four forms of short-term psychotherapy. *Psychol Psychother Theory Res Pract* 76, 133–144. doi: 10.1348/147608303765951177

Moldovan, R., Radu, M., Baban, A., and Dumitrascu, D. L. (2015). Evolution of Psychosomatic Diagnosis in DSM. Historical Perspectives and New Development for Internists. *Rom J Intern Med* 53, 25–30. doi: 10.1515/rjim-2015-0003

Mousavi Asl, E., Mahaki, B., Khanjani, S., and Mohammadian, Y. (2020). The Assessment of Alexithymia Across Positive and Negative Emotions: The Psychometric Properties of the Iranian Version of the Perth Alexithymia Questionnaire. *Iran J Psychiatry Behav Sci* 14, e102317. doi: 10.5812/ijpbs.102317

Pandey, R., Saxena, P., and Dubey, A. (2011). Emotion regulation difficulties in alexithymia and mental health. *Eur J Psychol* 7, 604–623.

Passardi, S., Peyk, P., Rufer, M., Wingenbach, T. S., and Pfaltz, M. C. (2019). Facial mimicry, facial emotion recognition and alexithymia in post-traumatic stress disorder. *Behav Res Ther* 122, 103436. doi: 10.1016/j.brat.2019.103436

Preece, D. A., Becerra, R., Boyes, M. E., Northcott, C., McGillivray, L., and Hasking, P. A. (2020). Do self-report measures of alexithymia measure alexithymia or general psychological distress? A factor analytic examination across five samples. *Pers Individ Dif* 155, 109721. doi: 10.1016/j.paid.2019.109721

Preece, D., Becerra, R., Allan, A., Robinson, K., and Dandy, J. (2017). Establishing the theoretical components of alexithymia via factor analysis: Introduction and validation of the attention-appraisal model of alexithymia. *Pers Individ Dif* 119, 341–352. doi: 10.1016/j.paid.2017.08.003

Preece, D., Becerra, R., Robinson, K., and Dandy, J. (2018a). Assessing alexithymia: psychometric properties and factorial invariance of the 20-item Toronto alexithymia scale in nonclinical and psychiatric samples. *J Psychopathol Behav Assess* 40, 276–287. doi: 10.1007/s10862-017-9634-6

Preece, D., Becerra, R., Robinson, K., Dandy, J., and Allan, A. (2018b). The psychometric assessment of alexithymia: Development and validation of the Perth Alexithymia Questionnaire. *Pers Individ Dif* 132, 32–44. doi: 10.1016/j.paid.2018.05.011

Rudell, K., and Diefenbach, M. A. (2008). Current issues and new directions in psychology and health: culture and health psychology. Why health psychologists should care about culture. *Psychol Heal* 23, 387–390. doi: 10.1080/08870440701864983

Saarijärvi, S., Salminen, J., and Toikka, T. (2001). Alexithymia and depression: a 1-year follow-up study in outpatients with major depression. *J Psychosom Res* 51, 729–733.

Säkkinen, P., Kaltiala-Heino, R., Ranta, K., Haataja, R., and Joukamaa, M. (2007). Psychometric properties of the 20-item Toronto Alexithymia Scale and prevalence of alexithymia in a Finnish adolescent population. *Psychosomatics* 48, 154–161. doi: 10.1176/appi.psy.48.2.154

Schumacker, R. E., and Lomax, R. G. (2016). *A beginner's guide to structural equation modeling*. New York, NY: Routledge.

Sifneos, P. E. (1973). The prevalence of 'alexithymic'characteristics in psychosomatic patients. *Psychother Psychosom* 22, 255–262. doi: 10.1159/000286529

Swart, M., Kortekaas, R., and Aleman, A. (2009). Dealing with feelings: characterization of trait alexithymia on emotion regulation strategies and cognitive-emotional processing. *PloS one* 4:e5751. doi: 10.1371/journal.pone.0005751

Taylor, G. J., Bagby, R. M., and Parker, J. D. (2003). The 20-Item Toronto Alexithymia Scale: IV. Reliability and factorial validity in different languages and cultures. *J Psychosom Res* 55, 277–283.

Vorst, H. C., and Bermond, B. (2001). Validity and reliability of the Bermond–Vorst alexithymia questionnaire. *Pers Individ Dif* 30, 413–434. doi: 10.1016/s0191-8869(00)00033-7

Wang, Y. P., and Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. *Braz J Psychiatry* 35, 416–431. doi: 10.1590/1516-4446-2012-1048

Watters, C. A., Taylor, G. J., Quilty, L. C., and Bagby, R. M. (2016). An examination of the topology and measurement of the alexithymia construct using network analysis. *J Per Assess* 98, 649–659. doi: 10.1080/00223891.2016.1172077

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 28 April 2021
doi: 10.3389/fpsyg.2021.656862

The Development of a Multidimensional Inventory for the Assessment of Mental Pain (FESSTE 30)

Karin Flenreiss-Frankl[1], *Jürgen Fuchshuber*[2] *and Human Friedrich Unterrainer*[2,3,4]*

[1] Department of Social Sciences, University of Nicosia, Nicosia, Cyprus, [2] Center for Integrative Addiction Research (CIAR), Grüner Kreis Society, Vienna, Austria, [3] Department of Religious Studies, University of Vienna, Vienna, Austria, [4] University Clinic for Psychiatry and Psychotherapeutic Medicine, Medical University Graz, Graz, Austria

Background: Although the term "mental pain" is often the subject of expert opinions regarding claims for damages, there is still no standardized questionnaire in the German-speaking area to operationalize this concept. Therefore, the aim of this work is the development and validation of a self-assessment measurement for psychological pain after traumatic events (FESSTE).

Methods: *A first version of the questionnaire was applied on a sample of the German speaking general population (N = 425; 88% female). After performing an item analysis and exploratory factor analysis, the questionnaire was shortened and tested on a second German speaking general population sample (N = 619; 89% female).* Finally, the newly developed questionnaire was related to the extent of traumatization (measured with a uniquely designed trauma checklist attached to the FESSTE) and already established instruments for the assessment of psychiatric symptom burden, which included the Brief Symptom Inventory-18 (BSI-18) and the Post-traumatic-Stress-Scale (PTSS-10).

Results: The final version of the FESSTE consists of a total of 30 items and covers the subscales "Somatization," "Depression," "Intrusive Memories," "Dissociation" and "Anxiety," and a total scale "Mental Pain." Based on the confirmatory factor analysis, it is assumed that the latent factor structure of the FESSTE can be best described as a bifactor-model. The final version shows a satisfactory model fit, high internal consistencies, and strong positive correlations with the BSI-18 and PTSS-10, as well as the extent of traumatic experiences.

Discussion: The FESSTE enables an operationalization of mental pain comprising five subscales and one total scale. What is more, the trauma checklist attached to the FESSTE allows for the standardized assessment of potentially traumatic experiences and the corresponding extent of these experiences. The results indicate that the FESSTE is a reliable and valid self-assessment procedure for mental pain, which is suitable for use in research and in expert practice.

Keywords: trauma, mental pain, questionnaire development, validation, factor analysis

OPEN ACCESS

Edited by:
Antonio De Padua Serafim, Universidade de São Paulo, Brazil

Reviewed by:
Andrea Svicher, University of Florence, Italy
Francisco Javier Cano-García, Sevilla University, Spain

***Correspondence:**
Human Friedrich Unterrainer
human.unterrainer@univie.ac.at

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *21 January 2021*
Accepted: *29 March 2021*
Published: *28 April 2021*

Citation:
Flenreiss-Frankl K, Fuchshuber J and Unterrainer HF (2021) The Development of a Multidimensional Inventory for the Assessment of Mental Pain (FESSTE 30). Front. Psychol. 12:656862. doi: 10.3389/fpsyg.2021.656862

INTRODUCTION

The Austrian judicative is increasingly aware of the phenomenon of mental pain. In correspondence to this, mental pain is often framed as a psychotraumatic state of suffering with the extent of a disease (Barolin et al., 1994). Psychosomatic as well as psychopathological suffering caused by bodily harm—even without demonstrable physical injury consequences—justify a claim to pain compensation if the psychological impairment is deemed to require treatment or at least can be diagnosed by a doctor. However, mere agitation or outrage about damage, does not qualify as mental pain (Danzl et al., 2019).

Based on Holczabek's (1976) classification of physical pain, which is commonly referenced in Austrian legal opinions regarding the assessment of compensations, Laubichler (2002) defined three categories of mental pain: (1) Mild mental pain which is distinguished by a slight impairment of the ability to work, since mild mental pain only occurs intermittently and incidentally. This means, for example, that depression, anxiety and fears can be adequately dealt with by the affected subject. (2) Moderate pain is characterized by a balance of the ability to carry out activities in a professional or other sense and the inability to do so. Hence, moderate mental pain implies that activities are possible but significantly impaired. (3) Finally, in the case of severe mental pain, the subject is helpless against the pain and unable to detach himself from it. Therefore, the implementation of useful activities is inhibited, as the mental pain is so paramount that they occupy the entire space of consciousness.

Historically, there have been several attempts to define mental pain. From a psychoanalytic perspective, Freud's "Studies on Hysteria" already discussed the phenomena of mental pain (Freud and Breuer, 1895/2000). In this work he related the avoidance of the perception of mental pain caused by traumatic events to the development of hysterical symptoms. In his later writings "Mourning and Melancholia" (Freud, 1924), as well as "Inhibitions, Symptoms and Anxiety" (Freud, 1926/1975), he attributes mental pain to the specific feelings of sadness after traumatic experiences of loss. According to Freud (1924, 1926/1975) mental pain is functionally in the service of the detachment of libidinal energy from the lost love object. He sees this in contrast to physical pain, which in his view is characterized by a deduction of narcissistic libido from his own body parts.

Later, Bakan (1968) similarly emphasized the important role of social loss experiences in the development of mental pain. According to its definition, mental pain denotes the awareness of a disturbance in the human tendency toward a state of wholeness and social well-being. Furthermore, Frankl (1963) associates psychological pain primarily with the feeling of an agonizing inner emptiness. For him, this is the result of a loss of meaningfulness in one's own life and can therefore only be modulated by adjusting values that enable the individual to give his life a new meaning. Another definition can be found in the writings of Sandler (1962) and Joffe and Sandler (1965). In their conceptualization, mental pain is the result of a discrepancy between the ideal self and the perception of the real self. These self-representations are composed of mental images that the individual has of oneself and one's social role.

Moreover, Baumeister's (1990) theory on suicide as an escape from oneself poses a similar approach. For him, mental pain is caused by a strongly pronounced self-perception of inadequacies, which in turn results from negative self-attributions with regard to one's own failures. Specifically, this means that if the ideal self and the actual perception of the actual self are far apart and this difference is ascribed to one's own failures. The subject perceives this as mental pain. Hence, for Baumeister, the underlying effect of mental pain is disappointment in oneself. This contrasts with the Herman's (1992) and Janoff-Bulman's (1992) conceptualization, who define mental pain as the perception of negative changes within the self. Similar to Freud (1926/1975, 1966) and Bakan (1968) they assume that these are essentially related to the experience of trauma and loss of attachment figures.

In the course of his work on suicidality, Shneidman (1977, 1993, 1998) dealt extensively with the phenomenon of mental pain, which he calls "psychache." He assumes that mental pain is fed by the frustration of essential needs of the individual, such as being loved, need for control and security, protection of self-image, avoidance of shame and the need for understanding. If these basic needs are not sufficiently met, a mixture of different negative emotional states such as guilt, shame, loss, despair, loneliness, grief, hopelessness, and anger occurs (Shneidman, 1993). In this context, Shneidman emphasizes that this variety of negative emotions can lead to a generalized and almost unbearable psychological pain. If this pain reaches an intensity which is too high and it is not possible to predict any positive change for the individual in the future, it often necessitates an attempt to evade this state through suicide.

Bion (1970/2013) introduced a differentiation between the concept of mental pain and suffering. He argued that the emergence of mental pain is linked to traumatic experiences described as "beta elements." These beta elements are experienced as overwhelming and are not accessible for the patient's "alpha function." Therefore, they fail to be contained and symbolized and are ultimately expressed as mental pain. In contrast, suffering is associated with negative experiences which can be contained within the mental apparatus. Hence, the emotional pain caused by these events can be processed into "alpha elements" which are suitable for mental elaboration and can be further processed by the patient (Bion, 1970/2013; Fleming, 2006).

Against this psychoanalytic background, Shneidman (1999) also made the first attempt at a questionnaire-based operationalization of the mental pain concept with the original *Psychological Pain Assessment Scale* (PPAS; Shneidman, 1999), which has never been empirically validated. However, on the basis of Shneidman's work, the *Psychache Scale* (Holden et al., 2001) was developed which assesses the intensity of mental pain based by means of 13 items, which are answered on a five-point Likert scale. Empirical findings show a high internal consistency of the one-factorial scale and confirm the construct validity of Shneidman's concept, particularly with regard to the assumed connections with depression, hopelessness, psychiatric symptoms, and increased suicidality (Holden et al., 2001; Mills et al., 2005). However, a German version of this questionnaire is not yet available.

Another approach to research into psychological pain was developed by Bolger (1999). Her definition takes as a starting point a qualitative analysis of narrative descriptions of mental pain in traumatized individuals. Based on these evaluations, she understands mental pain as a torn self ("brokenness of the self"). This conflict is made up of various determinants and includes a feeling of hurt, separation from loved ones, loss of self-esteem, loss of control, and fear. However, Bolger did not develop a standardized measuring instrument for the operationalization of mental pain, as well.

Also on the basis of a qualitative content analysis, but of narratives from psychiatric inpatients and randomly selected healthy test persons (see Orbach and Mikulincer, 2002), Orbach et al. (2003b) developed a questionnaire for the assessment of mental pain. In correspondence to this, Orbach et al. (2003a,b) define mental pain as a broad spectrum of subjective perception of negative changes within the self and its functions. This perception is accompanied by intense negative feelings and is therefore often experienced as torture. The *Orbach and Mikulincer Mental Pain Scale* (OMMP; Orbach et al., 2003b) has nine factors that are operationalized with 44 items that are answered on a five-point Likert scale. The operationalized scales are detailed as: *Irreversibility, Loss of Control, Narcissist Wounds, Emotional Flooding, Freezing, Self-estrangement, Confusion, Social Distancing*, and *Emptiness*. In its validation study the OMMP showed satisfying psychometric criteria (Orbach et al., 2003b). In addition, the results of Orbach et al. (2003a) suggest significant correlations between OMMP and increased suicidality and hopelessness, which is in line with the conceptual framework by Frankl (1963), Baumeister (1990), and Shneidman (1998). These relationships were also highlighted in a recently published systematic review by Verrocchio et al. (2016). Again, no translation has been carried out for the German-speaking area for the OMMP.

In the legal and expert context, mental pain is generally defined as "unbill" or "hardship," which means as an immaterial damage, which justifies corresponding compensation (Barolin et al., 1994). In expert practice, mental pain is often equated with mental illness. However, the diagnosis of post-traumatic stress disorder (PTSD) as defined by ICD-10 (World Health Organization, 1992) is not a prerequisite for classifying mental pain. It is assumed that, in addition to PTSD, symptoms such as a stress reaction, depression, anxiety disorder or adjustment disorder also cause mental pain. The presence of a mental pain or a psychotraumatic state of suffering of disease value should only be assumed, however, if there is a traumatic event, which must have a clearly demonstrable connection with the psychological state (Barolin et al., 1994).

Construction of the First Version of the Questionnaire

As a first step in the present study, a qualitative content analysis (Mayring, 2010) of 18 court opinions was carried out. Based on this the scales for the first empirical examination of the FESSTE were determined. The analysis resulted in 28 categories, which were grouped into the superordinate categories "Vegetative Symptoms/Somatization," "Treatment," "Compensation Behavior," "Vulnerability," "Anxiety Symptoms," "Trauma Disorder," "Depression," and "Mood Disorder." These categories were used in the next step as starting points for the construction of the individual scales. Due to the lack of meaningfulness, the categories "vulnerability" and "mental disorders" were excluded from the further scale construction. The category "treatment" is covered descriptively in the questionnaire through several separate questions but was not included in the calculation of the mental pain total value.

For the remaining categories—depending on the scope of the theoretical concepts—between 3 and 23 items were formulated. The selection and formulation of the items was based on the developed categories, the ICD-10 (World Health Organization, 1992) and on already established questionnaires for the individual concepts, such as the Symptom-Check-List-90 (SCL-90; Derogatis, 1975), and the Essener Trauma-Inventar (ETI; Tagay et al., 2007). In addition, the items were formulated in the form of statements. Particular emphasis was placed on adhering to specific formulation rules in order to increase comprehensibility and subsequently, the measurement accuracy (Bühner, 2011). These rules include avoiding double negations, using short, unambiguous sentences and avoiding foreign words or technical terms.

On the basis of empirical results that suggest that scaling with a response format of five to seven levels maximizes the reliability of the measuring instrument (Bühner, 2011) a five-point Likert scale was used to assess the individual items. "Not at all (0)—very strong (4)" was chosen as the response format.

Moreover, at the beginning of the questionnaire, a list of potentially traumatic events is given, which is based on the results of epidemiological data from Perkonigg et al. (2000) and the recent trauma definition of the DSM-5 (American Psychiatric Association, 2013), encompassing events connected to actual or menaced death, serious personal injury or sexual violence. The individual items can be answered with "yes" or "no" with regard to the question of whether they happened to the test participant. The category "Yes" is divided into the subcategories "Personal" and "As a witness." The given answers are evaluated with specific point values ("yes—personal" = 2; "yes—as a witness" = 1; no = 0), which can be summed up to assess the parameter "extent of traumatic experience."

DESIGN AND METHODS

Sample and Procedure

The sample from the general population examined in the trial phase consisted of 425 German-speaking participants (88% female; age: 18–76 years). In the validation phase, a sample of 619 participants (90% female; age: 18–72 years) was examined. The participants were recruited through advertisements in public forums and social networks (e.g., Facebook). The data was collected via the online study platform LimeSurvey. Informed consent was obtained from all subjects before they filled in the questionnaires. The survey consisted of various demographic questions (e.g., gender, age, education, and psychiatric diagnoses) as well as the standardized test procedures described below. The participants did not receive any compensation. The sole inclusion criterion for participating in the online survey was an age over 18. Participants remained completely anonymous during and after the period of study participation. The study was carried out in accordance with the Helsinki Declaration.

Measurement Dimensions

Trial Phase

A total of 77 items were formulated to cover each of the five initial categories "Vegetative Symptoms / Somatization," "Compensatory Behavior," "Anxiety Symptoms," "Trauma Disorder," and "Depression," which were obtained from the qualitative content analysis of court reports.

The queried list of potentially traumatic experiences includes in detail:

(1) Torture; (2) Stay in war zone; (3) Serious accident, fire, or explosion; (4) Natural disaster; (5) Serious illness or injury; (6) Sudden or unexpected death of a close relative or an important person; (7) Displacement and Migration; (8) Imprisonment; (9) Neglect; (10) Sexual violence by a stranger as an adult; (11) Sexual violence by a person from the family or circle of acquaintances; (12) Sexual violence as a child or adolescent by a stranger; (13) Sexual violence as a child or adolescent by a person from the family or circle of friends; (14) Violent attack by a stranger; (15) Violent attack by a person belonging to the family or circle of acquaintances. Furthermore, "Other stressful life events" were assessed as a descriptive dimension which was excluded from the "extent of traumatic experience" score.

Psychometric Assessment

In order to evaluate aspects of convergent validity, the following psychometric measurement tools were included in the trial phase:

Brief Symptom Inventory (BSI-18; Derogatis, 2001; German version: Spitzer et al., 2011). The BSI-18 is a short version of the SCL-90-R (Derogatis, 1975). This self-report measurement is used to assess psychological distress within the last 7 days and consists of 18 items. The questionnaire includes the subscales "Depression," "Anxiety," and "Somatization," as well as an overall scale for assessing the general burden of symptoms, which is called "Global Severity Index" (GSI). The individual items are answered using a five-point Likert scale, which ranges from 0 ("not at all") to 4 ("very much"). Reliability for the BSI-18 ranged from excellent to acceptable with Cronbach's $\alpha = 0.93$ (GSI) – 0.79 (Somatization).

The *Post-traumatische-Stress-Scale-10* (PTSS-10; Holen et al., 1990; German version: Maercker, 2003), is a screening instrument for the assessment of PTSD symptoms. It comprises 10 items, which are answered in the current version on a seven-point Likert scale (from 0 = "never" to 6 = "always"). Subjects were asked about the extent of typical PTSD symptoms within the last week. The questionnaire covers the following areas: (1) "Sleep Problems"; (2) "Nightmares"; (3) "Depression"; (4) "Jumpiness"; (5) "Withdrawal Tendencies"; (6) "Irritability"; (7) "Mood Swings"; (8) "Feelings of Guilt"; (9) "Anxiety," and (10) "Muscle Tension." The reliability of this scale was excellent with Cronbach's $\alpha = 0.90$ ($N = 425$).

Validation Phase

Based on the results of the item characteristic analysis, exploratory factor and reliability analysis, the revised version of the questionnaire was given to a new sample of the general population. For this aim, the questionnaire was shortened to overall 52 items, operationalizing the scales "Somatization" (9 items), "Anxiety" (12 items), "Depression" (14 items), "Intrusive Memories" (11 items), and "Dissociation" (6 items).

Statistical Analysis and Analysis Strategy

The statistical analysis was conducted via SPSS 25.0 and AMOS 24.0. SPSS was used for data management and the calculations of descriptive statistics, reliabilities, exploratory factor analysis, bivariate correlations, ANOVAs and multivariate hierarchical regression analysis. The confirmatory factor analysis was estimated with AMOS. The total sample of 1,044 participants is divided into two groups (A = 425; B = 619). In group A, an exploratory factor analysis (varimax rotation, analysis of the 5-factorial solution) and the calculation of aspects of convergent validity were carried out. A confirmatory factor analysis was performed on the data from group B. Descriptive statistics and reliability of the final FESSTE were assessed based on the total sample.

With regard to the confirmatory factor analysis, the following fit indices were accepted as indicators for an acceptable model fit (Kline, 2015): (a) Comparative Fit Index (*CFI*) >0.90; (b) Tucker-Lewis Index (*TLI*) >0.90; (c) Normed Fit Index (NFI) >0.90; (d) Square Root Error of Approximation (*RMSEA*) <0.08 with the upper limit of the 90% confidence interval <1; (e) a χ^2/df <3. The model fit of competing models was assessed by comparing the AIC values. The model with the lowest AIC value was preferred. A ΔAIC >2 was seen as an indication of a significant difference between the models (Cheung and Rensvold, 2002; Jovanović, 2015). In the case of a poor model fit, the items with the lowest power of discrimination with regard to the assigned factors were removed from the model. This process was repeated until the model had a satisfactory fit. In accordance with Kline (2015), all indicators were logarithmically transformed for the confirmatory factor analysis, due to the lack of a normal distribution of the individual items (Kline, 2015).

RESULTS

Sample Characteristics

The descriptive sample characteristics for the trial and validation phases are detailed in **Table 1**. The mean age of the participants of the trial phase (sample A; $N = 425$) and the validation phase (sample B; $N = 619$) was 35 years (A: SD = 13 years; B: SD = 11 years). Three hundred and seventy-four (88%) of the participants in sample A were female, compared to 555 (90%) participants in group B. Three hundred and eighty-one (A: 90%) and 600 (B: 96%), respectively, stated at least one lifetime traumatic experience or stressful life event. Most participants' highest educational qualification was a qualification for higher education

TABLE 1 | Sample characteristics (testing and validation phase).

Sample	Exploration phase	Validation phase
Overall	$N = 425$	$N = 619$
Gender	$N = 374$ Female (88%)/ $N = 44$ Male (22%)	$N = 555$ Female (90%)/ $N = 64$ Male (10%)
Age	M = 35 (*SD* = 13 Jahre)	M = 35 (*SD* = 11 Jahre)
At least one traumatic life event	$N = 383$ Yes (90%)/ $N = 42$ No (10%)	$N = 600$ Yes (96%)/ $N = 19$ No (4%)
Highest finished education	$N = 161$ General qualification for university entrance (38%)/ $N = 103$ Master University degree (24%)/ $N = 67$ Bachelor University degree (16%)/ $N = 53$ Apprenticeship (12%)/ $N = 21$ Phd (5%)/ $N = 17$ High School (4%)/ $N = 3$ None (1%)	$N = 185$ General qualification for university entrance (30%)/ $N = 140$ Master University degree (22%)/ $N = 121$ Bachelor University degree (19%)/ $N = 119$ Apprenticeship (19%)/ $N = 10$ Phd (2%)/ $N = 39$ High School (6%)/ $N = 5$ None (1%)
Occupation	$N = 210$ in employment (50%)/ $N = 128$ in education (30%)/ $N = 65$ Unemployed (15%)/ $N = 22$ in Pension (5%)	$N = 365$ in employment (59%)/ $N = 141$ in education (22%)/ $N = 81$ Unemployed (13%)/ $N = 32$ in Pension (5%)
Relationship status	$N = 156$ Single (32%)/ $N = 134$ in relationship (32%)/ $N = 84$ Married (20%)/ $N = 45$ Divorced (10%)/ $N = 6$ Widowed (1%)	$N = 201$ Single (32%)/ $N = 222$ in relationship (36%)/ $N = 118$ Married (19%)/ $N = 69$ Divorced (11%)/ $N = 9$ Widowed (1%)
Nationality	$N = 353$ Austrian (83%)/ $N = 40$ German (9%)/ $N = 32$ Other (8%)	$N = 532$ Austrian (86%)/ $N = 59$ German (10%)/ $N = 28$ Other (5%)
Psychiatric diagnosis	$N = 347$ No (81%)/ $N = 78$ Yes (19%)	$N = 542$ No (88%)/ $N = 77$ Yes (12%)

(A: $n = 161$; 38%; B: $n = 185$; 30%). Furthermore, the majority was in employment during the time of the study (A: $n = 210$; 50%; B: $n = 365$; 59%). Concerning the relationship status, the greater part of sample A was single ($n = 156$; 32%), which is contrasted by sample B which had a majority of participants in a relationship ($n = 134$; 36%). Regarding nationality most subjects in both groups were Austrian (A: $n = 353$; 83.1%; B: $n = 532$; 86%). A significant proportion of probands reported some form of medication due to health problems related to a traumatic incident (A: $n = 120$; 28%; B: $n = 228$; 37%) Finally, 78 (20%) participants in group A declared to be diagnosed with a psychiatric disorder, compared to 77 (12%) in group B.

Descriptive Analysis of the Extent of Traumatization

As detailed in **Table 2**, the most commonly reported traumatic event was the sudden death of a close relative or an important caregiver (A: $n = 212$; B: $n = 328$). The personal experience of violent attacks by people from family or acquaintances (A: $n = 107$; B: $n = 208$), and serious illness or injury (A: $n = 105$; B: $n = 191$) were also frequently stated. A large number of participants also reported some form of sexual abuse. In this context, "sexual violence by a person from family or friends as a child or adolescent" was mentioned most frequently in both groups (A: $n = 76$; B: $n = 127$). A substantial proportion of participants also reported the experience of other stressful life events (e.g., mobbing, being left by a partner, stalking, and loss of a job) not listed in the questionnaire (A: $n = 130$; B: $n = 219$). Furthermore, results of the reliability analysis regarding the "extent of traumatic experience" total score was within an acceptable range with Cronbach's $\alpha = 0.63$ ($N = 1{,}044$).

Exploratory Analysis of the FESSTE Factor Structure

The exploratory factor analysis was carried out on the basis of the main component analysis with a VARIMAX rotation and was a priori set to five factors. The five factors solution explained 56.96% of the total variance. In the rotated factor solution, the first factor "depressive symptoms" was comprised of 14 items with factor loadings ranging from 0.78 to 0.42, showed an eigenvalue of 8.43 and explained 16.20% of the variance. The further course of the components can be given with eigenvalues of 6.19 (11.90%) for "Dissociation" which included six items (factor loadings: 0.72 – 0.44); "Fear" which showed an eigenvalue of 5.61 with 10.79% explained variance and was comprised of 12 items with factor loadings ranging 0.77 – 0.45; "Intrusive Memories" exhibiting an eigenvalue of 5.47 with 10.52% explained variance and 11 Items (factor loadings ranging from 0.72 to 0.49); and finally, "Somatization" which had an eigenvalue of 3.92 with 7.54% explained variance and 9 items ranging from 0.70 to 0.46 regarding their factor loadings.

Analysis of the Latent Factor Structure of the FESSTE

As shown in **Table 3**, no model of the original 52-item version of the FESSTE showed a satisfactory fit (all models: CFI <0.90;

TABLE 2 | Sample characteristics regarding experienced trauma.

Sample	Trial phase (N = 381)				Validation phase (N = 600)			
Type of traumatization	**Personal**		**As witness**		**Personal**		**As witness**	
Torture	11	3%	13	3%	25	4%	11	2%
Stay in war zone	12	3%	9	2%	12	2%	12	2%
Serious accident, fire or explosion	58	15%	61	16%	92	15%	82	14%
Natural disaster	13	3%	28	7%	44	7%	34	6%
Serious illness or injury	105	28%	148	39%	191	32%	187	31%
Sudden/unexpected death of a close relative or an important person	212	56%	54	14%	328	55%	82	14%
Flight and migration	14	4%	42	11%	10	2%	34	6%
Captivity	10	3%	8	2%	11	2%	11	2%
Neglect	67	18%	49	13%	115	19%	51	9%
Sexual violence by a stranger as an adult	50	13%	10	3%	91	15%	22	4%
Sexual violence by a person from the family or circle of acquaintances as an adult	66	17%	16	4%	116	19%	13	2%
Sexual violence by a stranger as a child or adolescent	51	13%	9	2%	86	14%	10	2%
Sexual violence by a person from the family or circle of acquaintances as a child or adolescent	75	20%	14	4%	138	23%	14	2%
Violent attack by a stranger	76	20%	55	14%	127	21%	60	10%
Violent attack by a member of the family or circle of acquaintances	107	28%	24	6%	208	35%	36	6%
Other traumatic events	130	34%	14	4%	213	36%	12	2%

TABLE 3 | Results of the confirmatory factor analysis of several FESSTE models.

Modell	$\chi^2(df)$	χ^2/df	RMSEA (90% CI)	CFI	NFI	TLI	AIC
52-Item Version							
One-factor	8737.39 (1,280)	6.83	0.098 (0.096–0.100)	0.58	0.54	0.56	8933.39
Five-factor	4897.27 (1,264)	3.87	0.069 (0.067–0.071)	0.80	0.74	0.76	4897.27
One-factor of higher order	4948.04 (1,269)	3.90	0.069 (0.067–0.071)	0.79	0.74	0.78	5166.04
Bifactor	4656.32 (1,228)	3.79	0.069 (0.066–0.070)	0.80	0.76	0.79	4956.32
30-Item Version							
One-factor	3465.33 (405)	8.56	0.112 (0.108–0.115)	0.68	0.65	0.65	3465.33
Five-factor	1331.76 (395)	3.37	0.063 (0.059–0.066)	0.90	0.87	0.89	1471.76
One-factor of higher order	1359.48 (400)	3.40	0.063 (0.059–0.067)	0.90	0.86	0.89	1489.49
Bifactor	1031.03 (375)	2.75	0.054 (0.050–0.058)	0.93	0.90	0.92	1211.03

N = 619.

TLI <0.90; NFI <0.90; $\chi^2/df > 3$). Subsequently, the number of items for the respective factors was step wise reduced, in consideration of the selectivity of the separate items. Overall, this approach led to a substantial increase in the model fit. In the 30-item version, the individual subscales of the FESSTE showed a generally acceptable model fit. The most promising parameters were achieved with the bifactor model of the 30-item version of the FESSTE. In this model, all global fit indices were satisfactory: RMSEA = 0.05 (90% CI: 0.05, 0.06); TLI = 0.92; CFI = 0.93; NFI = 0.90; AIC = 1211.03.

As shown in **Figure 1**, this model showed a strong overall factor "mental pain," which is significantly associated with all items (β = 0.38–0.80; all $p < 0.001$). Overall, the loadings of the sub-factors on the respective items were somewhat less pronounced. Regarding the subscales, "fear" showed significant associations in the range between β = 0.28–0.68 (all $p < 0.001$). The regression coefficients on the scale "Intrusive Memories" scale ranged between β = 0.19–0.69 (all $p < 0.001$). "Depression" was associated with the assigned items within the range β = 0.11–0.50 (p <0.01–0.001). The associations between the items and the latent factor "dissociation" were between β = 0.29–0.55 (all $p < 0.001$). Finally, the "somatization" scale showed associations in the range β = 0.27–0.56 (all $p < 0.001$).

Correlation With Psychiatric Symptoms and Psychometric Properties of the FESSTE

As can be seen in **Table 4**, the overall scale showed excellent internal consistency (Cronbach's α = 0.95). Furthermore, the subscales also showed satisfactory internal consistencies (Cronbach's α = 0.82–0.91). **Table 5** also details the correlations of the subscales with the overall scale and the intercorrelations between the subscales. With regard to the correlation patterns, it should be noted that the subscales of the FESSTE were intercorrelated less with one another than with the overall scale

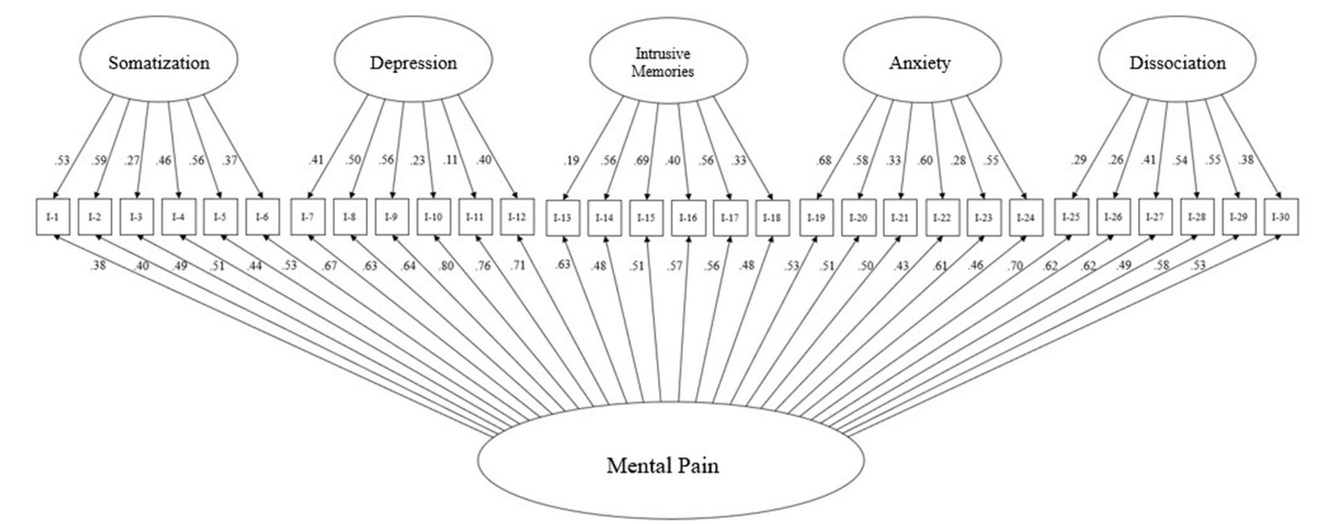

FIGURE 1 | Bifactor model of the 30-item version of the Questionnaire for the Assessment of Mental Pain after Traumatic Events (FESSTE 30) ($N = 619$).

TABLE 4 | Correlation between the FESSTE scales and indicators of symptom severity.

Variable	FESSTE 30 total score	FESSTE 30 somatization	FESSTE 30 depression	FESSTE 30 anxiety	FESSTE 30 dissociation	FESSTE 30 intrusive memories
Extent of traumatic experiences	0.52*	0.48*	0.42*	0.42*	0.45*	0.42*
BSI depressiveness	0.78*	0.50*	0.90*	0.55*	0.65*	0.58*
BSI anxiety	0.83*	0.63*	0.72*	0.79*	0.63*	0.68*
BSI somatization	0.73*	0.80*	0.54*	0.59*	0.52*	0.61*
GSI (BSI-18)	0.91*	0.72*	0.86*	0.74*	0.70*	0.72*
PTSS-10	0.88*	0.67*	0.81*	0.70*	0.69*	0.75*

*$N = 425$. *$p < 0.001$; GSI = BSI-18 total score.*

TABLE 5 | Internal consistency and intercorrelation of the FESSTE 30 scales.

Variable	1	2	3	4	5	6
1. Somatization	–					
2. Intrusive memories	0.52*	–				
3. Anxiety	0.53*	0.58*	–			
4. Depression	0.52*	0.62*	0.55*	–		
5. Dissociation	0.47*	0.57*	0.54*	0.70*	–	
6. FESSTE total score	0.74*	0.83*	0.78*	0.86*	0.81*	–
Cronbach's α	0.82	0.86	0.87	0.91	0.86	0.95

*$N = 1{,}044$; *$p < 0.001$; M, mean; SD, standard deviation.*

value, which can be interpreted as a preliminary confirmation of the derived factor structure.

As shown in **Table 5**, the subscales and the overall scale of the FESSTE show medium to high positive correlations with the external criteria ($r = 0.50$ -0.91; all $p < 0.001$) and with the reported extent of experienced traumatic experiences ($r = 0.42$ -0.52; all $p < 0.001$; see Cohen, 1992). Strong correlations were observed between the FESSTE total score and the BSI-18 ($r = 0.91$; $p < 0.001$) and the PTSS-10 ($r = 0.88$; $p < 0.001$).

Regarding the examination of the distribution characteristics which was based on the Kolmogorov-Smirnov adaptation test for normal distribution and the analysis of the skewness and kurtosis, it can be assumed that none of the scales exhibited a normal distribution (see **Table 6**). In a sample of the general

TABLE 6 | Questionnaire for the Assessment of Mental Pain after Traumatic Events (FESSTE 30): mean, standard deviation, and distribution characteristics.

Variable	M	SD	Min	Max	Skew	Kurtosis	Test for normal distribution*	
							z	*p*
1. Somatization	6.47	5.42	0	24	0.84	−0.08	4.23	0.00
2. Intrusive memories	9.02	6.76	0	24	0.46	−0.82	3.28	0.00
3. Anxiety	4.85	5.58	0	24	1.38	1.19	6.22	0.00
4. Depression	8.87	6.98	0	24	0.48	−0.90	3.78	0.00
5. Dissociation	5.71	5.79	0	24	1.04	0.27	5.24	0.00
6. FESSTE total score	34.94	24.67	0	114	0.54	−0.47	2.75	0.00

*N = 1,044; M, mean; SD, standard deviation. *Kolmogorov-Smirnov-test.*

population, the subscales and the overall scale show a clearly skewed right-hand distribution. With regard to kurtosis, the scales "Somatization," "Intrusive Memories," "Depression," and the overall scale show a flattened distribution, while the scales "Fear" and "Dissociation" show a sharp distribution.

Influence of Age and Gender on Mental Pain

With regard to age effects on the FESSTE scales, small negative correlations between age and the subscales "trauma-related disorder" ($r = -0.07$; $p < 0.05$), "anxiety" ($r = -0.06$; $p < 0.05$) and dissociation ($r = -0.10$; $p < 0.05$) were observed. Moreover, there is a small positive correlation between age and the extent of traumatic experiences ($r = -0.07$; $p < 0.05$). No significant correlations were found between the age of the subjects and the FESSTE total scale ($r = -0.04$; $p = 0.21$) and the Subscale "Depressive Symptoms" ($r = -0.04$; $p = 0.17$).

In the overall scale and all subscales, female subjects showed significantly higher values ($F = 22.96$–57.63; all $p < 0.001$; $\eta^2 = 0.02$–0.05). Women also reported a slightly higher extent of traumatic experiences ($F = 5.36$; $p < 0.05$; $\eta^2 = 0.01$).

Comparison of the Prediction of the Extent of Traumatic Experiences Between the FESSTE, BSI-18, and PTSS-10

A comparison regarding the prediction of the extent of traumatic experiences between the established measuring instruments PTSS-10 and BSI-10, as well as the newly developed FESSTE was carried out with a multivariate hierarchical regression analysis. For this aim, sex and age was entered as a control variable at step 1 ($R^2 = 0.04$; $\Delta F = 8.93$; $p < 0.001$). In step 2 the PTSS-10 total score and the BSI-18 total score were added ($R^2 = 0.25$; $\Delta F = 60.25$; $p < 0.001$). Finally, step 3 included the FESSTE total score ($R^2 = 0.30$; $\Delta F = 28.75$; $p < 0.001$). The results suggested that the FESSTE enables a significantly higher explanation of variance than the two comparison instruments ($\Delta R = 0.05$; $\beta = 0.59$; $p < 0.001$; see **Table 7**).

TABLE 7 | Hierarchical multiple regression to predict the extent of traumatic experience controlled for gender and age.

	Variable	R^2	ΔR^2	ΔF	β
Step 1		0.04	0.04	8.93***	
	Gender				−0.18***
	Age				0.12*
Step 2		0.25	0.21	60.25***	
	Gender				−0.04
	Alter				0.14**
	PTSS-10				0.33***
	GSI				0.17*
Step 3		0.30	0.05	28.75***	
	Gender				−0.05
	Age				0.15***
	PTSS-10				0.07
	GSI				0.09
	FESSTE Total score				0.59***

*N = 425; Criteria = Extent of traumatic experience; Gender was coded as: Female = 0; Male = 1; GSI = BSI-18 total score; ***p < 0.001; **p < 0.01; and *p < 0.05.*

DISCUSSION

The aim of the present work was to develop and validate a self-report measurement for the operationalization of mental pain, since this concept is of great importance in expert practice (Laubichler, 2002), yet to date no standardized questionnaires for the German speaking area exists which operationalizes this construct.

On the basis of qualitative evaluations of judicial reports, 8 factors were originally assumed, which could be reduced to an overall factor and five sub-factors in the course of the test development. With regard to the results of the factor-analytical evaluations, a bifactor model with a central factor "mental pain" and the five domain-specific factors "Somatization," "Depression," "Anxiety," "Intrusive Memories," and "Dissociation" could be established. The originally assumed dimensions "Compensation Behavior" and "Vulnerability" could not be confirmed. The aspect of "treatment," which is important for expert practice, is assessed descriptively in the questionnaire, but is not included in the calculation of the total score. In the final version of the questionnaire, this total value of "mental pain" is formed by adding up the five sub-scales to which six items are assigned.

Originally, 77 items were constructed with regard to theoretical considerations and observations from the qualitative content analysis. In consecutive steps, 43 items were eliminated based on the results of the item analysis and analysis of the factor structure. Despite this item reduction, the values of the internal consistency, which are in a consistently satisfactory range for the final version with 30 items, indicate a high level of reliability (Bühner, 2011). In order to determine the stability of the FESSTE, however, further studies to determine the retest reliability must be carried out.

The preliminary test version obtained through exploratory factor analysis with 52 items with five factors could not be confirmed in a subsequent confirmatory factor analysis. However, a further item reduction based on the item-total correlation analysis resulted in a 30-item version of the questionnaire with a satisfactory model fit. In a direct comparison between several conceivable models, the bifactor model proved to be the latent structure with the best fit with regard to the empirically obtained data. It can therefore be assumed that the FESSTE is made up of an overall scale of "mental pain" and five residual factors that can be interpreted as subordinate factors (Chen et al., 2006). In correspondence to this, one might understand the high intercorrelations between the individual sub-factors (Cohen, 1992), resembling the correlation pattern of the BSI sub scales (Franke et al., 2011), which according to more recent findings exhibits a latent bifactor structure, as well (Thomas, 2012; Urbán et al., 2014).

Furthermore, the results of this study indicate a substantial positive correlation between the FESSTE and the applied measures for post-traumatic stress disorder and general psychiatric symptoms. These results suggest a high degree of convergent validity (Bühner, 2011). In particular, the overlap between the global scale of the BSI-18 and the FESSTE is in an area that raises the question of a differentiation between the general psychiatric distress measured with the BSI and the concept of mental pain. It should be noted that there is an increasing trend in recent psychiatric literature which instead of a purely categorical separation between mental illnesses, recognizes both broad, unspecific factors and domain-specific factors with regard to the latent structure of psychiatric phenomena (Watson, 2005; Tackett et al., 2008; Goodkind et al., 2015). In correspondence to this, it can be assumed that the mental pain measured with the FESSTE has a strong conceptual overlap with the global symptom severity measured with the BSI. The further calculations with a hierarchical multiple regression show, however, that the FESSTE is superior to both the PTSS-10 and the BSI-10 in predicting the extent of traumatic experiences. Hence, the data obtained in this study suggested the conclusion that the FESSTE is a more suitable measuring instrument for expert work. To further investigate the validity of FESSTE, future studies could investigate connections with the Orbach and Mikulincer Mental Pain Scale (OMMP, Orbach et al., 2003b), as well as increased suicidality (see Shneidman, 1998).

The total score of the FESSTE shows no significant age effects and the FESSTE subscales also only show negligible correlations with age, if at all. On the other hand, there are slight to medium gender differences in the answers to the FESSTE, which suggest a higher degree of mental pain in women. This finding is in line with epidemiological studies that show a higher prevalence of internalizing disorders, such as depression and PTSD in women (Kendler et al., 2003; Kessler, 2003; Bangasser and Valentino, 2014), while men show a higher probability of externalizing disorders—such as addictions (Grant et al., 2015). The latter are not recorded in the current version of FESSTE. An originally created scale that would have recorded externalizing "compensatory behavior" was removed due to inadequate psychometric properties. However, since traumatic addiction development and other externalizing disorders play a role in the assessment of mental pain, albeit in rarer cases, an extension of the FESSTE in this regard could be useful.

The current version of FESSTE has a trauma checklist oriented toward the results of Perkonigg et al. (2000) and the recently published DSM-5 (American Psychiatric Association, 2013). This enables experts to make an economic and structured assessment of the extent of the experienced traumatic events. The results imply that the traumatization experienced was closely related to the self-assessed mental pain. However, new research results show that this connection is substantially mediated by the structural integrity of personality structure (Fuchshuber et al., 2019; Flenreiss-Frankl et al., 2021). The mental pain triggered by traumatization should therefore be understood within the framework of an interplay of traumatic events and resilience or vulnerability that vary between individuals. What is more, it has to be noted that the extent of traumatic experiences scale showed only low reliability in this study. Regarding the wide variety of possible traumatic events, this circumstance might be expected. While the scale might be useful for practitioners concerned with forensic opinions due to its descriptive value, researcher who employ this scale into their studies should be conscious of its limited internal consistency and hence, interpret its score with caution.

Limitations and Research Perspectives

A limitation of the validation study of the FESSTE results from the imbalance of the gender ratio within the sample. Comparatively few men took part in the examinations, which prevents the questionnaire from being standardized according to gender. On a similar note, the investigated sample shows an imbalance toward a relatively high proportion of participants with high education levels, which further inhibits the standardization of the scale based on the general population. In order to establish more reliable norm values and investigate for factorial invariance of the latent factor structure, it is therefore crucial that future studies are undertaken which explore more representative population samples.

Moreover, the FESSTE, like comparable self-assessment procedures, is prone to deliberate falsification. The use of suitable control procedures for socially desirable response styles could represent a possible addition to FESSTE. However, these instruments are only of limited effectiveness (van de Mortel, 2008). Hence, expert instinct and experience seem to play a particularly important role in assessing mental pain.

Furthermore, the ability of FESSTE to assess the time of mental pain after trauma is very limited. For this purpose, the development of a standardized interview is planned, which is created based on the questionnaire and enables an assessment of the respective symptom groups using temporal categories.

In addition to the creation of a standardized interview based on the results of the FESSTE, the development of a short version of the FESSTE is planned, which could be suitable to further sharpen the psychometric properties and the economic validity of the FESSTE.

In this study, which was based on data of the general population, the FESSTE scales did not show a normal distribution. However, comparable measuring instruments like the BSI showed similar results (Gilbar and Ben-Zur, 2002; Spitzer et al., 2011). In contrast, a normal distribution of the FESSTE scales is to be expected in clinical populations. It is therefore necessary to investigate patient groups in future studies concerned with norm data of the FESSTE.

Finally, the results might be influenced by the mode of data collection. While recent evaluations of the validity of online surveys underscored that they are comparable to more traditional surveys (Wiersma, 2013), it is yet to be investigated if the results of the present study can be extended to the paper-pencil version of the FESSTE.

CONCLUSION

The aim of the work was to develop a questionnaire able to assess mental pain after traumatic experiences. With the FESSTE, a corresponding instrument is now available which enables an operationalization of mental pain on the basis of five subscales and one total scale. What is more, the trauma checklist attached to the FESSTE allows for the standardized assessment of potentially traumatic experiences and the corresponding extent of these experiences. In particular, the questionnaire is—with the exception of the extent of traumatic event scale—characterized by very high internal consistencies regarding its subscales and overall scale. Moreover, the results of the confirmatory factor analysis confirm the satisfactory structural validity of the questionnaire construction. The strong associations with relevant external criteria—such as psychiatric symptoms and the extent of traumatic experiences—indicate the high criterion validity of this new procedure. Finally, the FESSTE can be used without high economic expense, which makes it a useful instrument for approaching mental pain after traumatic events. While the limitations of this study discussed above imply the need for further research regarding standardization and factorial invariance, in summary, the results of this first examination indicate that the FESSTE is a reliable and valid self-assessment procedure, suitable for the use in research and in expert practice.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics board, University of Nicosia. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KF-F, JF, and HU developed the study and conducted the data analysis. JF conducted the data collection. KF-F and JF wrote the first draft of the manuscript. HU proofread the manuscript and made some critical comments. All authors read the final version of the manuscript and gave their consent for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.656862/full#supplementary-material

REFERENCES

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®).* Washington, DC: American Psychiatric Pub. doi: 10.1176/appi.books.9780890425596

Bakan, D. (1968). *Disease, Pain, and Sacrifice: Toward a Psychology of Suffering.* Chicago, IL: University of Chicago Press.

Bangasser, D. A., and Valentino, R. J. (2014). Sex differences in stress-related psychiatric disorders: neurobiological perspectives. *Front. Neuroendocrinol.* 35, 303–319. doi: 10.1016/j.yfrne.2014.03.008

Barolin, G. S., Griebnitz, E., Mitterauer, B., Quatember, R., Scherzer, E., and Spiel, W. (1994). Die Begutachtung sogenannter seelischer Schmerzen. *Der Sachverständige* 2, 131–133.

Baumeister, R. F. (1990). Suicide as escape from self. *Psychol. Rev.* 97, 90–113. doi: 10.1037/0033-295X.97.1.90

Bion, W. R. (1970/2013). *Attention and Interpretation: A Scientific Approach to Insight in Psycho-Analysis and Groups.* Abingdon: Routledge.

Bolger, E. (1999). Grounded theory analysis of emotional pain. *Psychother. Res.* 9, 342–362. doi: 10.1093/ptr/9.3.342

Bühner, M. (2011). *Einführung in die Test-und Fragebogenkonstruktion.* Hallbergmoos: Pearson Deutschland GmbH.

Chen, F. F., West, S. G., and Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivar. Behav. Res.* 41, 189–225. doi: 10.1207/s15327906mbr4102_5

Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Mode* 9, 233–255. doi: 10.1207/S15328007SEM0902_5

Cohen, J. (1992). Statistical power analysis. *Curr. Dir. Psychol. Sci.* 1, 98–101. doi: 10.1111/1467-8721.ep10768783

Danzl, K.-H., Gutiérrez-Lobos, K., and Müller, O. F. (2019). *Das Schmerzengeld: in medizinischer und juristischer Sicht;[mit Entscheidungsteil].* Vienna: Manz.

Derogatis, L. R. (1975). *SCL-90-R: Symptom Checklist-90-R: Administration, Scoring, and Procedures Manual.* London: NCS Pearson.

Derogatis, L. R. (2001). *Brief Symptom Inventory 18.* Baltimore, MD: Johns Hopkins University. doi: 10.1037/t07502-000

Fleming, M. (2006). Distinction between mental pain and psychic suffering as separate entities in the patient's experience. *Int. Forum Psychoanal.* 15, 195–200. doi: 10.1080/08037060500522754

Flenreiss-Frankl, K., Unterrainer, H. F., and Fuchshuber, J. (2021). Die Beziehung zwischen Trauma, Traumafolgen und Persönlichkeitsstruktur: Eine Mediationsanalyse. *Zeitschrift für Psychosomatische Medizin und Psychotherapie,* 67. doi: 10.13109/zptm.2021.67.oa3

Franke, G. H., Ankerhold, A., Haase, M., Jäger, S., Tögel, C., Ulrich, C., et al. (2011). Der Einsatz des brief symptom inventory 18 (BSI-18) bei Psychotherapiepatienten. *Psychother. Psychosom. Med. Psychol.* 61, 82–86. doi: 10.1055/s-0030-1270518

Frankl, V. (1963). *Viktor Frankl.* Munich: Bertelsmann.

Freud, S. (1924). Mourning and melancholia. *Psychoanal. Rev.* 11:77.

Freud, S. (1926/1975). *Hemmung, Symptom und Angst.* Frankfurt: Fischer.

Freud, S. (1966). "Mourning and melancholia," in *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume I (1886-1899): Pre-Psycho-Analytic Publications and Unpublished Drafts* (London: Hogart Press), 244–258.

Freud, S., and Breuer, J. (1895/2000). *Studies on Hysteria.* New York, NY: Basic Books.

Fuchshuber, J., Hiebler-Ragger, M., Kresse, A., Kapfhammer, H. P., and Unterrainer, H. F. (2019). The influence of attachment styles and personality organization on emotional functioning after childhood trauma. *Front. Psychiatry* 10:643. doi: 10.3389/fpsyt.2019.00643

Gilbar, O., and Ben-Zur, H. (2002). Adult Israeli community norms for the brief symptom inventory (BSI). *Int. J. Stress Manag.* 9, 1–10. doi: 10.1023/A:1013097816238

Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-Hagata, L. B., et al. (2015). Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* 72, 305–315. doi: 10.1001/jamapsychiatry.2014.2206

Grant, B. F., Goldstein, R. B., Saha, T. D., Chou, S. P., Jung, J., Zhang, H., et al. (2015). Epidemiology of DSM-5 alcohol use disorder. Results from the National Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA Psychiatry* 72, 757–766. doi: 10.1001/jamapsychiatry.2015.0584

Herman, J. L. (1992). *Trauma and Recovery.* New York, NY: Basic Books.

Holczabek, W. (1976). Gerichtsmedizinische Grundlagen der Schmerzbestimmung. *Forschung und Praxis d. Begutachtung* 12, 11–44.

Holden, R. R., Mehta, K., Cunningham, E. J., and McLeod, L. D. (2001). Development and preliminary validation of a scale of psychache. *Can. J. Behav. Sci.* 33, 224–232. doi: 10.1037/h0087144

Holen, A., Sund, A., and Weisaeth, L. (1990). *The Alexander Kielland Disaster March 27th 1980: Psychological Reactions among the Survivors.* Oslo: Division of Disaster Psychiatry, University of Oslo.

Janoff-Bulman, R. (1992). *Toward a New Psychology of Trauma:* New York, NY: Free Press.

Joffe, W. G., and Sandler, J. (1965). Notes on pain, depression, and individuation. *Psychoanal. Study Child.* 20, 394–424. doi: 10.1080/00797308.1965.11823243

Jovanović, V. (2015). Structural validity of the Mental Health Continuum-Short Form: the bifactor model of emotional, social and psychological well-being. *Personal. Individ. Differ.* 75, 154–159. doi: 10.1016/j.paid.2014.11.026

Kendler, K. S., Prescott, C. A., Myers, J., and Neale, M. C. (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch. Gen. Psychiatry* 60, 929–937. doi: 10.1001/archpsyc.60.9.929

Kessler, R. C. (2003). Epidemiology of women and depression. *J. Affect. Disord.* 74, 5–13. doi: 10.1016/S0165-0327(02)00426-3

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling.* New York, NY: Guilford publications.

Laubichler, W. (2002). "Schmerzengeld aus neurologisch-psychiatrischer Sicht," in *Das ärztliche Gutachten*, ed H. Emberger (Vienna: ÖÄK-Verlag), 267–278.

Maercker, A. (2003). "Posttraumatische-Stress-Skala-10 (PTSS-10)," in *Angstdiagnostik–Grundlagen und Testverfahren*, eds J. Hoyer and J. Margraf (Berlin: Springer), 401–403.

Mayring, P. (2010). "Qualitative inhaltsanalyse," in *Handbuch qualitative Forschung in der Psychologie*, eds G. Mey and Mruck (Wiesbaden: Springer), 601–613. doi: 10.1007/978-3-531-92052-8_42

Mills, J. F., Green, K., and Reddon, J. R. (2005). An evaluation of the Psychache Scale on an offender population. *Suicide Life Threat. Behav.* 35, 570–580. doi: 10.1521/suli.2005.35.5.570

Orbach, I., and Mikulincer, M. (2002). *Mental Pain and Pathology*. Unpublished Study, Department of Psychology, Bar-Ilan University, Ramat-Gan.

Orbach, I., Mikulincer, M., Gilboa-Schechtman, E., and Sirota, P. (2003a). Mental pain and its relationship to suicidality and life meaning. *Suicide Life Threat. Behav.* 33, 231–241. doi: 10.1521/suli.33.3.231.23213

Orbach, I., Mikulincer, M., Sirota, P., and Gilboa-Schechtman, E. (2003b). Mental pain: a multidimensional operationalization and definition. *Suicide Life Threat. Behav.* 33, 219–230. doi: 10.1521/suli.33.3.219.23219

Perkonigg, A., Kessler, R. C., Storz, S., and Wittchen, H.-U. (2000). Traumatic events and post-traumatic stress disorder in the community. Prevalence, risk factors and comorbidity. *Acta Psychiatr. Scand.* 101, 46–59. doi: 10.1034/j.1600-0447.2000.101001046.x

Sandler, J. (1962). Psychology and psychoanalysis. *Br. J. Med. Psychol.* 35, 91–100. doi: 10.1111/j.2044-8341.1962.tb00507.x

Shneidman, E. (1977). *Definition of Suicide.* Lanham: Jason Aronson, Incorporated.

Shneidman, E. S. (1993). Commentary: suicide as psychache. *J. Nervous Mental Dis.* 181, 145–147. doi: 10.1097/00005053-199303000-00001

Shneidman, E. S. (1998). *The Suicidal Mind.* Oxford: Oxford University Press.

Shneidman, E. S. (1999). The psychological pain assessment scale. *Suicide Life Threat. Behav.* 29, 287–294.

Spitzer, C., Hammer, S., Löwe, B., Grabe, H. J., Barnow, S., Rose, M., et al. (2011). Die Kurzform des Brief Symptom Inventory (BSI-18). Erste Befunde zu den psychometrischen Kennwerten der deutschen Version. *Fortschr. Neurol. Psychiatr.* 79, 517–523. doi: 10.1055/s-0031-128 1602

Tackett, J. L., Quilty, L. C., Sellbom, M., Rector, N. A., and Bagby, R. M. (2008). Additional evidence for a quantitative hierarchical model of mood and anxiety disorders for DSM-V: the context of personality structure. *J. Abnorm. Psychol.* 117, 812–825. doi: 10.1037/a0013795

Tagay, S., Erim, Y., Stoelk, B., Möllering, A., Mewes, R., and Senf, W. (2007). Das Essener Trauma-Inventar (ETI)–Ein Screeninginstrument zur Identifikation traumatischer Ereignisse und posttraumatischer Störungen. *ZPPM* 1, 75–89. doi: 10.1055/s-2006-934318

Thomas, M. L. (2012). Rewards of bridging the divide between measurement and clinical theory: demonstration of a bifactor model for the Brief Symptom Inventory. *Psychol. Assess.* 24, 101–113. doi: 10.1037/a0024712

Urbán, R., Kun, B., Farkas, J., Paksi, B., Kökönyei, G., Unoka, Z., et al. (2014). Bifactor structural model of symptom checklists: SCL-90-R and Brief Symptom Inventory (BSI) in a non-clinical community sample. *Psychiatry Res.* 216, 146–154. doi: 10.1016/j.psychres.2014.01.027

van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Aust. J. Adv. Nurs.* 25, 40–48.

Verrocchio, M. C., Carrozzino, D., Marchetti, D., Andreasson, K., Fulcheri, M., and Bech, P. (2016). Mental pain and suicide: a systematic review of the literature. *Front. Psychiatry* 7:108. doi: 10.3389/fpsyt.2016.00108

Watson, D. (2005). Rethinking the mood and anxiety disorders: a quantitative hierarchical model for DSM-V. *J. Abnorm. Psychol.* 114, 522–536. doi: 10.1037/0021-843X.114.4.522

Wiersma, W. (2013). The validity of surveys: online and offline. *Oxford Inter. Inst.* 18, 321–340.

World Health Organization (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines.* Geneva: World Health Organization.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 03 June 2021
doi: 10.3389/fpsyg.2021.658989

The Development and Validation of the Psychological Needs of Cancer Patients Scale

Yao Chen[1†], Fangyan Lin[2†], Bo Wang[3], Yung-lung Tang[2], Jun Li[4*] and Lin Xiong[3*]*

[1] Radiation Therapy Center, Chongqing University Cancer Hospital, Chongqing, China, [2] Faculty of Psychology, Southwest University, Chongqing, China, [3] Department of Psychology, Chongqing University Cancer Hospital, Chongqing, China, [4] Department of Urologic Oncology Surgery, Chongqing University Cancer Hospital, Chongqing, China

OPEN ACCESS

Edited by:
Victor Zaia,
Faculdade de Medicina do ABC,
Brazil

Reviewed by:
Laura Galiana,
University of Valencia, Spain
Maryam Rassouli,
Shahid Beheshti University of Medical
Sciences, Iran

***Correspondence:**
Yung-Lung Tang
tyl57525@126.com
Jun Li
herrleej@foxmail.com
Lin Xiong
739700897@qq.com

†These authors have contributed
equally to this work

Specialty section:
This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 26 January 2021
Accepted: 19 April 2021
Published: 03 June 2021

Citation:
Chen Y, Lin F, Wang B, Tang Y-l,
Li J and Xiong L (2021) The
Development and Validation of the
Psychological Needs of Cancer
Patients Scale.
Front. Psychol. 12:658989.
doi: 10.3389/fpsyg.2021.658989

In the present research, the Psychological Needs of Cancer Patients Scale (PNCPS) was developed and validated. Based on Group 1 (400 cancer patients), the exploratory factor analysis identified a 23-item scale with six factors: value and esteem (five items, i.e., reconsider the meaning and purpose of life), independence and control (six items, i.e., private space), mental car (three items, i.e., vent negative emotions), disease care (three items, i.e., acquire knowledge about disease), belonging and companionship (three items, i.e., spend more time at home), and security (three items, i.e., living conditions be better). The structure identified with Group 1 was further tested, based on Group 2 (199 cancer patients), for reliability and validity. The results showed that PNCPS has a clear factor structure and good psychometric characteristics. By taking into account the cultural background of Chinese patients, this scale will advance the study of the psychological needs of those with malignant tumors and thus has a certain reference value for other countries.

Keywords: psychological needs, cancer patients, scale development, palliative care, hospice care, mental health

INTRODUCTION

According to World Health Statistics 2020, cancer caused 9 million deaths in 2016 (WHO, 2020). In 2018, cancer was responsible for an estimated 9.6 million deaths. And according to the latest statistics from China's National Cancer Center, the estimated incidence of malignant tumors nationwide was 285.83 per 100,000 in 2015. Malignant tumor patients form a large patient group. Importantly, the negative impact of cancer on patients is not only reflected in their physical symptoms but also reflected through social adjustment disorders and psychological discomfort (Peng et al., 2019). Although the reported prevalence of depression varies significantly (Massie, 2004; Pirl, 2004), the majority of rates for major depressive disorder fall between 10 and 25% of patients with malignant cancer tumors, while for probable cases of depression, the majority of reports fall into the 7–21% range (Pirl, 2004). Similarly, the reported prevalence of anxiety problems in cancer patients is higher than in others without cancer (Stark and House, 2000). Additionally, the fear of recurrence and post-traumatic stress disorder are two common psychological concerns among cancer survivors (Claire et al., 1997; Kangas et al., 2002; Strong et al., 2008).

Therefore, it is likely that many psychological needs of malignant tumor patients are not met. Need refers to a lack and imbalance in the human tissue system. In the present research, we aimed to develop a reliable and valid measure of the psychological needs of malignant tumor patients, appropriate for use in the Chinese cultural context.

The development of such a measure is important. First of all, investigating and meeting the psychological needs of patients with malignant tumors are conducive to improving their quality of life as part of palliative care and hospice care. With the development and progress of modern medicine, the survival time for malignant tumor patients is prolonged, and the treatment of cancer has become a long and painful process. From the moment of diagnosis, palliative care improves the quality of life of patients and their families facing physical, psychological, social, and spiritual challenges related to life-threatening disease and is an essential component of cancer care (WHO, 2021). An estimated 40 million people need palliative care each year, yet unfortunately, only about 14% of those in need worldwide are currently receiving it (WHO, 2020).

At the end of a patient's life, hospice care provides comprehensive and active whole-person care for the patient and his/her family members. By providing psychological care and comfort, eliminating the fear of death, and paying attention to the patient's comfort, the dying can face death calmly and leave safely (Chen, 2003). A survey and analysis of end-of-life care needs of patients with advanced cancer found that patients had high levels of physical (e.g., sleep and pain) and psychological (e.g., anxiety and depression) needs, and moderate levels of spiritual (e.g., death education) and social support (e.g., family contact) needs (Zhao et al., 2015). Treatment currently focuses on cure and therefore does not acknowledge the other but very real needs of the majority of patients who will die from their disease (Hilden et al., 2001). Under the cancer resource allocation scheme recommended by the WHO for developing countries, more resources should be devoted to pain control and end-of-life care for patients with advanced cancer than to cancer treatment itself (Coyne, 1997).

Investigating and meeting the psychological needs of patients with malignant tumors will help promote their mental health, as people with cancer often have mental health problems. Most current research focuses on reducing or preventing the suffering and negative reactions experienced by cancer survivors, thus promoting their mental health. For example, Strong et al. (2008) conducted a psychological intervention on cancer patients with depression and found that depression, anxiety, and fatigue in patients were improved. However, it is important to note that psychological health in cancer survivors is determined by both the presence or absence of distress as well as the presence or absence of a variety of positive psychological responses often subsumed under the concept of "posttraumatic growth" (Cordova et al., 2001; Cordova and Andrykowski, 2003; Stanton et al., 2006). However, very little research has looked at whether and, if so, how positive psychological responses could be fostered in cancer survivors. Research into the psychological needs of cancer patients could help compensate for this. In other words, while focusing on unmet needs, we can also focus on well-met needs and tap into the positive reactions of cancer survivors to promote their mental health.

A better understanding of the psychological needs of patients with malignant tumors can also contribute to improving their physical health. Both the occurrence and development of malignant tumors are closely related to the psychological state of patients. A large number of clinical studies and epidemiological investigations have found that some specific psychosocial factors, such as stress and depression, are risk factors for the occurrence and development of tumors. Thus, people with poor mental health are more likely to develop tumors, demonstrating that psychosocial factors are closely related to the occurrence, development, and prognosis of tumors (Price et al., 2001). Many researchers have found that psychological interventions for patients with malignant tumors can also have a positive impact on the patient's condition. The research of Guo et al. (2019) showed that a personalized psychological intervention can effectively improve anxiety, depression, and negative mood in patients with lung cancer and improve their quality of life, satisfaction, and medication compliance. In addition to the above factors, psychological needs also include the patient's need to understand medical information and medical services, which are directly related to treatment planning and treatment effects.

The satisfaction of psychological needs over the course of cancer treatment will affect treatment satisfaction and treatment effects (Walker et al., 2003). However, the current health-care systems worldwide are unable to meet the various psychological needs of patients (Walker et al., 2003; Marijnen et al., 2005; Cox et al., 2006), and patients are less satisfied with how their psychological needs are met by doctors than they are with the diagnosis and treatment of cancer.

In 2000, Billie Bonevski published the first questionnaire to assess the psychological needs of cancer patients (Bonevski et al., 2000). Since then, a range of tools have been developed to assess the needs of patients with cancer. The purpose and population of these studies vary a lot: some were proposed as specific to advanced stage of disease, clinical setting, or survivors, and some were targeted particular diagnoses (e.g., lung cancers) (Duke et al., 2003; Hodgkinson et al., 2007; Ahmed et al., 2011; Richards et al., 2011). Among then, a number of instruments were developed to assess multiple needs of general population, such as the Supportive Care Needs Survey (SCNS) and its short form (SCNS-SF34), the Cancer Rehabilitation Evaluation System (CARES), the Supportive Needs Screening Tool (SNST) (Bonevski et al., 2000; Boyes et al., 2009; Pigott et al., 2009; Johnsen et al., 2011). However, the assessment of the psychological needs of cancer patients in China started slightly later than in other countries; research in this field has mostly use translated and revised questionnaires developed in other countries. In particular, the SCNS-SF34 reportedly has good internal validity and reliability (Richardson et al., 2007; Carlson et al., 2012) and perform well in the Chinese population (Li et al., 2013; Choi et al., 2020). Whether it is the best fit for the Chinese population, however, remains uncertain. Localized and widely recognized reliable questionnaires have not been developed.

Thus, the aim of the present study was to construct the Psychological Needs of Cancer Patients Scale (PNCPS) with the Chinese population.

ITEM GENERATION

As mentioned earlier, need refers to a lack and imbalance in the human tissue system. According to different standards, needs can be divided into different categories. Among them, Maslow's hierarchy of needs theory (Maslow, 1943) is a representative theory of needs, and it has many implications for not only education and teaching but also the present study. Maslow's hierarchy of needs divides human needs into five categories from low to high (like a ladder), as follows (Maslow, 1943; Koltko-Rivera, 2006):

Physiological (survival) needs: Seeks to obtain the basic necessities of life, such as food, water, air, sleep, and sex. They are the most important and powerful of human needs.

Safety needs: Seeks security through order and law.

Belonging and love needs: Seeks affiliation with a group.

Esteem needs: Seeks esteem through recognition or achievement.

Self-actualization: Seeks fulfillment of personal potential.

Generally speaking, when the needs of a certain level are relatively satisfied, they will develop to a higher level, and the pursuit of a higher level of needs will become the driving force for behavior. Correspondingly, the need to obtain basic satisfaction of needs is no longer an incentive.

Based on Maslow's (1943) theory, open-ended questionnaires were conducted through semi-structured interviews completed by a representative sample of malignant tumor patients (see **Table 1**) to understand their psychological needs in depth and detail and to identify as many pertinent issues as possible from their perspective. In addition, four experts were invited for evaluation and consultation. This initial work was used to determine the dimensions and items of the preliminary questionnaire.

TABLE 1 | Characteristics of participants: open-ended questionnaire.

Variable		*n*/(*M* ± *SD*)
Gender	Male	12
	Female	12
Marital status	Married	23
	Widowed	1
Tumor stage	I	6
	II	6
	III	6
	IV	6
Department	Radiotherapy general ward	2
	Gynecological oncology	9
	Urinary oncology	8
	Thoracic oncology	5
Age		56.46 ± 11.28

Open-Ended Questionnaire

An open-ended questionnaire on the psychological needs of cancer patients was designed. The open-ended questionnaire mainly consists of five questions, which were corresponding with Maslow's five needs. For example, the question about physiological needs was "How is your sleep recently? And what other physiological needs do you think are important to you?"

To ensure the data quality of the on-site interviews, interviewers were postgraduates selected from psychology majors and with experience in conducting on-site interviews. Before the formal interview, the interviewers were trained in the interview process and questionnaire content. Discussions and exchanges on the coding consistency of various topics were conducted to ensure the consistency of attitudes and interview content among interviewers.

A total of 24 patients (see **Table 1**) representing both genders and a range of ages, stages, and types of cancer were selected for in-depth face-to-face interviews. Written informed consent from the patients or their guardians was obtained before the interview. Centering on the topic, the interview was carried out with questions started with "how," "what," "when," and "why." Patients were asked to describe in detail the problems they had met in each stage of the disease, and specifically, the content and degree of the psychological needs, under what circumstances did the needs arise, and how satisfied these psychological needs were. The duration of the interview depends on the patients' physical conditions and cooperation degree, and the duration of the interview for all patients ranges from 5 to 43 min. After the consent of the patients was obtained, the interviews were recorded and later coded.

In order to collect items as widely as possible, we extracted all the key words concerning "need" (whether satisfied or not) that appeared in the recordings and encoded them as sentences "I hope . . . ." Finally, an item pool containing 107 items was formed.

Expert Consultation

The opinions of clinicians and experts (including an associate professor of psychology and three postgraduates in psychology) were obtained to develop the preliminary questionnaire. After their review and discussion on all 107 items, (1) items with similar meaning or overlapping content should be deleted or merged; (2) items with abstract and vague meanings should be modified to facilitate thinking by patients; and (3) optimize the language of all projects for easy reading and clear understanding, and items were managed.

Preliminary Questionnaire Preparation

A total of 58 items were formed. After the initial questionnaire items were determined, a random number string was generated to reorder all items. All items started with "I hope . . . " and multiple-choice responses were anchored on a five-point scale (1 = completely inconsistent, 2 = less inconsistent, 3 = uncertain, 4 = more consistent, and 5 = completely consistent).

MATERIALS AND METHODS

Through open-ended questionnaire and expert consultation, we obtained the initial scale. Next, we obtained the patient's performance on the scale through questionnaire method. A series of analysis, including item screening, exploratory factor analysis (EFA), and confirmatory factor analysis (CFA), were performed to develop the PNCPS.

Participants and Sampling

The cluster sampling method was adopted, and Chongqing Cancer Hospital was selected as the survey site. Participants were randomly selected according to the following requirements: aged =18 years; pathological diagnosis of cancer; intact cognitive function and normal ability to express understanding; adequate physical condition to complete the questionnaire; and able to provide signed informed consent and participate voluntarily. A total of 698 participants completed the questionnaire, of whom 599 (323 females) met the data screening criteria specifying that the continuous selection of the same option did not exceed 28 questions (Curran, 2016).

To construct the scale, 400 patients (Fabrigar et al., 1999; Lloret et al., 2014) were randomly selected from all subjects as Group 1, including 213 females, aged between 18 and 84 years (M = 51.87; SD = 12.53). The remaining 199 patients constituted Group 2, including 110 females, aged between 18 and 87 years (M = 53.73; SD = 12.85), were used to confirm the structure.

Additionally, the whole sample, including 599 patients (323 females), aged between 18 and 87 years (M = 52.49; SD = 12.66), was used to acquire information about psychometric properties. But only 441 patients' scores of depression and anxiety were recorded in the hospital information system and were used for verification.

Measures

Psychological Needs of Cancer Patients Scale

The preliminary version of the scale included 58 items. Responses were made on a five-point scale (1 = completely inconsistent, 2 = less inconsistent, 3 = uncertain, 4 = more consistent, and 5 = completely consistent). The higher the score, the more pressing the patient's corresponding needs. As can be seen in the "Results and Discussion" section, the final version consisted of 23 items. Internal consistency (Cronbach's alpha) was 0.83 and 0.85 for the confirmatory and the validation samples, respectively.

Depression Self-Rating Scale

Severity of depression was assessed using this 20-item scale (Zung, 1965) with responses made, as follows: 1 = no or very little time; 2 = a small amount of time; 3 = a lot of time; and 4 = most or all of the time. Ten items were reverse-scored. We added the scores of the 20 items to get the total raw score and then multiplied this by 1.25 to obtain the standard score. According to Chinese norms, the cutoff value of the Depression Self-rating Scale (SDS) standard score is 53 points, and 53–62 points indicate mild depression, 63–72 points indicate moderate depression, and 73 points or more indicate severe depression. The Cronbach α in this study was 0.91.

Anxiety Self-Assessment Scale

Severity of anxiety was measured using this 20-item scale (Zung, 1971) with responses made as follows: 1 = no or very little time; 2 = a small amount of time; 3 = a lot of time; and 4 = most or all of the time. Five items were reverse-scored. We added the scores of the 20 items to get the total raw score and then multiplied this by 1.25 to obtain the standard score. According to Chinese norms, the cutoff value of the Anxiety Self-Assessment Scale (SAS) standard score is 50 points, and 50–59 points indicate mild depression, 60–69 points indicate moderate depression, and 69 points or more indicate severe depression. The Cronbach α in this study was 0.89.

Procedure

After patients were informed of the purpose of the study, written or verbal informed consent was obtained. Only participants who met the eligibility criteria were tested. The questionnaire was presented to each participant on a computer and during testing, and the testing personnel avoided the influence of irrelevant factors. All data were obtained between December 3, 2019, and April 5, 2020. The study was reviewed and approved by the Ethics Committee of Chongqing Cancer Hospital.

STATISTICS

Firstly, Group 1 was used to explore the scale structure, which consisted of two steps: (1) screening for more efficient items using based on classical test theory (CTT); and (2) EFA was conducted. Secondly, CFA was conducted with Group 2 to confirm the rationality of the structure explored in the previous step. Finally, sample 3 was used to analyze psychometric properties of the scale.

The EFA and CFA were conducted with M Plus 8.6, and other analyses were processed with SPSS 22.0.

RESULTS AND DISCUSSION

Item Screening

Internal Consistency Coefficient

We calculated the Cronbach α coefficient of the scale. If the value of the new Cronbach α coefficient do not decrease after deleting an item, then the relationship between this question and other questions in the scale is weak, and this item should be considered for exclusion. In this step, 12 items was eliminated: 11, 14, 16, 19, 24, 26, 30, 40, 50, 55, 56, and 57. After seven items were deleted, the Cronbach α of the 46 items was 0.88.

Corrected Item–Total Correlation

We considered items with low corrected item–total correlations as problematic. And seven items (3, 10, 18, 46, 47, 48, and 53) were excluded from subsequent analyses.

In addition, the averages of 39 items range between 3.14 and 4.76, and the skewness range between −2.24 and −0.13.

Exploratory Factor Analysis

Based on the results of Item Screening, 39 items were included in the EFA. Considering the non-normal distribution of item scores, EFA was run using the robust weighted least square mean and variance (WLSMV) estimation (Beauducel and Herzberg, 2006; Muthén and Muthén, 2013). Item responses were treated as categorical variables, and oblique rotations using the GEOMIN method were generated.

During the EFA process, items that load strongly (>0.35) onto factors were retained. Also, a model was accepted with at least three items for each dimension.

The results of EFA show that the eight-factor model fits well: χ^2 = 784.33, *df* = 457, comparative fit index (*CFI*) = 0.94, Tucker–Lewis index (*TLI*) = 0.91, root mean square error of approximation (*RMSEA*) = 0.04, and standardized root mean square residual (*SRMR*) = 0.05; but one of the factors only contains two items. Based on the eight-factor model, nine items (2, 4, 5, 7, 9, 20, 21, 22, and 25) were removed due to factor loads below 0.35. The remaining 30 items were transferred to the next EFA.

Using the same process, we repeated the factor analysis two more times to explore the best fitting latent structure. In the first round, three items (6, 8, and 34) were removed because the coefficient was below 0.35. In the second round, four items (1, 23, 36, and 38) were removed as the coefficient was below 0.35.

Ultimately, the six-factor model consisting of 23 items fits well: χ^2 = 290.38, *df* = 130, *CFI* = 0.96, *TLI* = 0.91, *RMSEA* = 0.06, and *SRMR* = 0.05, and each factor contains at least three items. Combined with the meaning of each item, the six factors were defined as follows: value and esteem, independence and control, mental care, disease care, belonging and companionship, and security. And the corresponding eigenvalues for sample correlation matrix were 5.61, 1.98, 1.81, 1.74, 1.43, and 1.28, respectively. Factor loads are shown in **Table 2**.

It is important to note that the factor load of item 49 onto Factor 5 is 1.06. Due to the specificity of the group of cancer patients, the distribution of the score is biased, which led to the result.

It should also be noted that there were still some obvious cross-loading associated with a few items. Item 43 not only loads on value and esteem but also loads high on disease care. Item 39 loads on mental care and on disease care.

Confirmatory Factor Analysis

To confirm the six-factor structural model constructed in Group 1, CFA was performed using the WLSMV estimation. For this six-factor model in Group 2, χ^2 = 459.17, *df* = 194, *CFI* = 0.88, *TLI* = 0.86, and *RMSEA* = 0.08. We modified the model by adding three correlations: items 15 and 17, items 42 and 44, and items 49 and 51. The results showed that the six-factor model constructed in Group 1 was consistent with the data in Group 2, χ^2 = 363.99, *df* = 191, *CFI* = 0.92, *TLI* = 0.90, and *RMSEA* = 0.07. The normalized coefficients for each path and factor loads in Group 2 are shown in **Figure 1**.

TABLE 2 | EFA: 23 items and corresponding factor loads.

Item	F1	F2	F3	F4	F5	F6	Content
15	**0.70**						I hope to do some outdoor activities like walking and playing ball
13	**0.66**	−0.14					I hope to know the diagnosis of my disease
12	**0.58**						I hope to reconsider the meaning and purpose of life
17	**0.54**						I hope there are fewer people who know about my health condition
43	**0.40**		0.19	0.32			I hope I can take part in social work and continue to make use of my value
29	0.17	**0.69**					I hope to stay away from death, from things that remind me of death
27	0.25	**0.65**					I hope to understand death, face death, and think about death
31		**0.63**					I hope it is convenient to walk out
33		**0.63**					I hope to have more private space
32		**0.47**	0.19			0.14	I hope to get help financially
28		**0.43**					I hope I can make my own decisions on whether to receive treatment and what treatment to receive
37			**0.79**				I hope I can vent sadness, fear, anger, or other negative emotions
39			**0.66**	0.30			I hope to know about changes in my condition and related information as soon as possible
35	−0.15	0.35	**0.53**				I hope I can feel useful to my family
44				**0.74**			I hope the pain of eating can be eased
42				**0.72**			I hope to acquire some knowledge about my disease
41	0.28	0.16	0.23	**0.39**			I hope to talk to a professional psychologist about my mental state
49					**1.06**		I hope I can live a fulfilling life
51					**0.49**		I hope I can spend more time at home
45				0.27	**0.42**	−0.16	I hope my family can stay with me often
52						**0.79**	I hope to be able to complete daily activities independently
54	0.22					**0.75**	I hope I can confide in others
58		0.20				**0.42**	I hope the living conditions could be better

EFA, exploratory factor analysis.
Bold values emphasize the loads of items on the corresponding factor, to distinguish from other cross-loads.

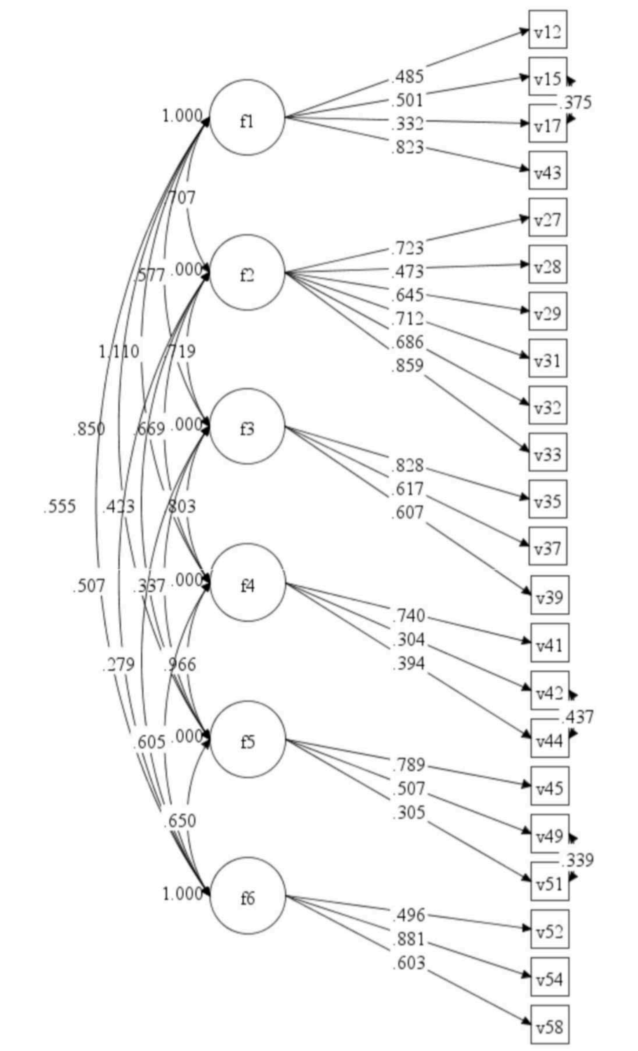

FIGURE 1 | The six-factor model and its standardized factor loading.

The four-item factor loads were below 0.40, which is essentially related to the Cronbach α coefficient.

Psychometric Properties of Psychological Needs of Cancer Patients Scale

Internal Consistency

Inter-correlations between subscales ranged from 0.20 to 0.53, all statistically significant at the 0.01 level (see **Table 3**).

In terms of internal consistency and reliability, the Cronbach α of the six subscales of value and esteem, independence and control, mental care, disease care, belonging and companionship, security, and the entire scale were 0.62 (five items), 0.71 (six items), 0.61 (three items), 0.42 (three items), 0.55 (three items), 0.49 (three items), and 0.84, respectively.

When the number of items is too small, it will violate the assumption of tau-equivalence, which underlies the Cronbach α, and will underestimate reliability (Kottner and Streiner, 2010; Tavakol and Dennick, 2011; Taber, 2017). In a sense, combined with the low-factor-loading, they indicate the multiple aspects of need and the relative independence of each factor. Given that the Cronbach α of the entire scale was 0.84, we consider it as a high degree of internal consistency.

Correlation With Anxiety Self-Assessment Scale and Depression Self-Rating Scale

Pearson correlations were conducted to explore the associations between PNCPS and the standard measures of depression and anxiety. The results showed that there was no significant correlation, but SAS was significantly correlated with value and esteem, independence and control, and PNCPS, and correlations were weak ($rs = 0.10$) (see **Table 3**).

Influence of Gender and Age on Psychological Needs of Cancer Patients Scale

An independent samples t-test showed that there were significant differences in psychological need scores between males and females. Specifically, males scored higher than females in value and esteem [$t(597) = 3.14$, $p = 0.00$ (two-tailed test)], independence and control [$t(596) = 2.32$, $p = 0.02$ (two-tailed test)], and PNCPS [$t(597) = 2.54$, $p = 0.01$ (two-tailed test) see **Table 4**].

There was also no correlation between age and psychological need scores (see **Table 3**), which indicates the invariance across age of the scale.

GENERAL DISCUSSION

We conducted this study to develop a reliable and effective scale to measure the psychological needs of patients with malignant tumors. Based on Maslow's (1943) hierarchy of needs theory, we conducted semi-structured interviews with a sample of representative malignant tumor patients to understand their psychological needs in depth and detail, and in doing so, we identified a large range of concerns. Experts in relevant fields also participated in evaluation and consultation. Based on this work, the dimensions and items of the preliminary

TABLE 3 | Correlations of the factors.

	F1	F2	F3	F4	F5	F6	PNCPS
F1	1						
F2	0.53**	1					
F3	0.28**	0.43**	1				
F4	0.40**	0.46**	0.31**	1			
F5	0.32**	0.32**	0.26**	0.28**	1		
F6	0.33**	0.38**	0.22**	0.32**	0.20**	1	
PNCPS	0.76**	0.85**	0.59**	0.66**	0.53**	0.58**	1
Age	0.06	0.03	0.02	–0.01	0.01	0.04	0.04
SAS	–0.10*	–0.10*	–0.07	–0.04	0.01	–0.07	–0.10*
SDS	0.02	–0.02	0.00	–0.04	0.00	–0.07	–0.02

*PNCPS, Psychological Needs of Cancer Patients Scale; SAS, Anxiety Self-Assessment Scale; SDS, Depression Self-Rating Scale. **Correlation is significant at the 0.01 level (two-tailed). *Correlation is significant at the 0.05 level (two-tailed).*

TABLE 4 | Influence of gender on PNCPS.

	Gender	*M*	*SD*	*t*	*df*	*p*
F1	Male	21.53	2.75	3.14**	597	0.002
	Female	20.79	3.02			
F2	Male	25.32	3.58	2.32*	596	0.021
	Female	24.60	4.02			
F3	Male	13.04	1.84	0.77	597	0.44
	Female	12.92	1.95			
F4	Male	12.88	1.88	0.64	597	0.525
	Female	12.78	1.94			
F5	Male	13.24	1.65	0.32	597	0.751
	Female	13.20	1.54			
F6	Male	13.18	1.73	1.96	597	0.051
	Female	12.88	2.04			
PNCPS	Male	99.19	9.53	2.54*	597	0.011
	Female	97.16	9.91			

*PNCPS, Psychological Needs of Cancer Patients Scale. **Correlation is significant at the 0.01 level (two-tailed). *Correlation is significant at the 0.05 level (two-tailed).*

questionnaire were preliminarily determined, and a final scale consisting of 23 items and six dimensions was constructed running with EFA. The results of the next CFA showed that the six-factor structure fitted the data well. In terms of reliability and validity, the Cronbach α showed that the scale as a whole has high internal validity. Additionally, the scale also showed invariance across age. Therefore, despite some minor flaws, we conclude that the current 23-item scale is a reliable, theory-based tool for measuring the psychological needs of patients with malignant tumors.

Four factors extracted from our research results (value and esteem, independence and control, belonging and companionship, and security) highlight the middle three needs of Maslow's hierarchy of needs, namely, the needs for safety, esteem, and belonging and love. There is only one item concerning basic physiological need, "I hope the pain of eating can be eased." This is in line with the situation of cancer patients. Faced with the direct threat of cancer, patients can do only very limited things, and other lower-level needs are more likely to be met.

Our research also identified two additional factors relating to the needs of cancer patients, namely, mental care and disease care. Similarly, in Zebrack's (2009) study, nearly all cancer patients (98.1%) expressed a need for information about the disease, treatment, and prognosis. A study of myeloma patients by Yogaparan et al. (2009) showed that more than two-thirds of older patients (aged over 60 years) needed to know about the life-prolonging effects of treatment (77%) and the side effects of treatment (67%), but more than 10% of patients thought they were given too little information about these two issues. According to Sollner et al. (2004), 31% of patients showed moderate-to-severe anxiety and depression in the early stages of treatment, and 41% strongly expressed the need for psychological support. This is both understandable and consistent with Maslow's hierarchy of needs theory.

In addition, there is the very interesting aspect of death. Our results show that cancer patients need to know about and face death. But in the context of Chinese culture, people's views on death have been influenced by Confucianism, Taoism, Buddhism, and other perspectives for a long time. In China, people always take a negative and veiled attitude toward death and do not mention death in words, considering it as a symbol of misfortune and fear (Ding, 2000). However, death is a problem that cancer patients, especially those at an advanced stage, have to face. This contradiction explains the necessity of hospice care and death education. Therefore, we suggest that we should pay attention to the care of the dying patients and implement death education, so that they can live without pain and die with dignity.

Despite a number of studies from the perspective of self-determination theory (Ryan and Deci, 2000), the satisfaction of basic psychological needs seems to predict a decrease in anxiety and depression (Wei et al., 2005; Yu et al., 2016). However, the same result was not found in this research. There was no significant correlation between psychological needs and depression and only a slight negative correlation between psychological needs and anxiety in patients with malignant tumors. One possible explanation is that different definitions of psychological needs and the tools used to measure them, as well as different groups of people, will lead to differing results. In addition, depression and anxiety related to cancer are different from general depression and anxiety. The occurrence of depression and anxiety is a complex process and not caused by one single factor. In addition to psychological factors, the development of depression and anxiety is also related to therapeutic drugs, hormones, and other chemicals (Wu and Xiao, 2011). Finally, the difference may also be caused by the unsatisfactory testing environment. Due to economic and time constraints, the questionnaire was not completed by patients individually under the supervision of professionals, so it was difficult to ensure that all patients responded carefully.

The development of a scale to assess the psychological needs of malignant tumor patients in China is of great significance. Only by understanding the content of patients' most urgent needs for help can we provide targeted help to cancer patients, make the optimal allocation of medical service resources, carry out mental health education and intervention, prevent possible psychological problems, improve quality of life, and achieve the simultaneous treatment of body and mind. It is of great practical significance to promote mental health and prolong the survival of patients with malignant tumors.

There are regional differences in the incidence levels of malignant tumors, and the psychological needs of patients with malignant tumors in different regions may also be different. Research in China in this field has not developed a localized, widely recognized, or reliable questionnaire. Participants in this study were from the largest cancer hospital in southwest China and may represent the whole of China to some extent. The PNCPS may become a widely used questionnaire as it takes the Chinese cultural background into account and fills existing gaps in this field. In this way, the PNCPS may contribute to a more accurate understanding of the psychological needs of Chinese, even the whole word's, malignant tumor patients and provide a basis for the formulation of targeted and effective measures.

This study also has some limitations. First, the results are based on data from patients in Chongqing Cancer Hospital, China, and the same results may not be obtained in other countries. Second, this was a cross-sectional survey, and the retest reliability of the scale has not been examined. Thus, the time stability of the scale remains to be confirmed. Third, as terminal cancer patients suffer great pain, it is difficult for them to support and cooperate with research, and terminal cancer patients were less represented in the sample, which may have some influence on the structure of the questionnaire. In conclusion, the validity of the PNCPS needs to be tested in more studies, and the scale needs to be revised and improved accordingly.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Chongqing Cancer Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Y-LT, LX, and JL led the design and implement of the study, including the literature search, analysis, and interpretation of the data. YC and FL led the drafting, writing, and revising of manuscript. All authors contributed to the questionnaire construction, elaboration, data collection, analysis, interpretation of data, and read and approved the final version of the work to be published and agreed to be accountable for all aspects of the work in ensuring that any question to the accuracy of the work is appropriately investigated and resolved.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.658989/full#supplementary-material

REFERENCES

Ahmed, K., Miskovic, D., Darzi, A., Athanasiou, T., and Hanna, G. B. (2011). Observational tools for assessment of procedural skills: a systematic review. *Am. J. Surg.* 202, 469–480. doi: 10.1016/j.amjsurg.2010.10.020

Beauducel, A., and Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Struct. Equat. Model.* 13, 186–203. doi: 10.1207/s15328007sem1302_2

Bonevski, B., Fisher, R. S., Girgis, A., Burton, L., Cook, P., and Boyes, A. (2000). Evaluation of an instrument to assess the needs of patients with cancer. *Cancer* 88, 217–225. doi: 10.1002/(sici)1097-0142(20000101)88:1<217::aid-cncr29>3.0.co;2-y

Boyes, A., Girgis, A., and Lecathelinais, C. (2009). Brief assessment of adult cancer patients' perceived needs: development and validation of the 34—item Supportive Care Needs Survey (SCNS—SF34). *J. Evaluat. Clin. Pract.* 15, 602–606. doi: 10.1111/j.1365-2753.2008.01057.x

Carlson, L. E., Waller, A., and Mitchell, A. J. (2012). Screening for distress and unmet needs in patients With cancer: review and recommendations. *J. Clin. Oncol.* 30, 1160–1177. doi: 10.1200/jco.2011.39.5509

Chen, A. P. (2003). Progress in hospice care for elderly patients. *Chin. J. Nurs.* 7, 65–67.

Choi, E., Liao, Q., Soong, I., Chan, K., Lee, C., Ng, A., et al. (2020). Measurement invariance across gender and age groups, validity and reliability of the Chinese version of the short-form supportive care needs survey questionnaire (SCNS-SF34). *Health Qual. Life Outcom.* 18:29.

Claire, L. J., Gerry, H., Rosaleen, D., and Mal, B. H. (1997). Fear of cancer recurrence – a literature review and proposed cognitive formulation to explain exacerbation of recurrence fears. *Psycho Oncol.* 6, 95–105. doi: 10.1002/(SICI)1099-1611(199706)6:2<95::AID-PON250<3.0.CO;2-B

Cordova, M. J., and Andrykowski, M. A. (2003). Responses to cancer diagnosis and treatment: posttraumatic stress and posttraumatic growth. *Semin. Clin. Neuropsych.* 8:286.

Cordova, M. J., Cunningham, L., Carlson, C. R., and Andrykowski, M. A. (2001). Posttraumatic growth following breast cancer: a controlled comparison study. *Health Psychol. Off. J. Divis. Health Psychol. Am. Psychol. Assoc.* 20, 176–185. doi: 10.1037/0278-6133.20.3.176

Cox, A., Jenkins, V., Catt, S., Langridge, C., and Fallowfield, L. (2006). Information needs and experiences: An audit of UK cancer patients. *Eur. J. Oncol. Nurs.* 10, 263–272. doi: 10.1016/j.ejon.2005.10.007

Coyne, P. J. (1997). International efforts in cancer pain relief. *Semin. Oncol. Nurs.* 13, 57–62. doi: 10.1016/S0749-2081(97)80051-4

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *J. Exp. Soc. Psychol.* 66, 4–19. doi: 10.1016/j.jesp.2015.07.006

Ding, Y. (2000). Ethical issues in the development of hospice care. *Chin. J. Nurs.* 1, 43–45.

Duke, J. M., Treloar, C. J., and Byles, J. E. (2003). Evaluation of an instrument to assess the needs of men diagnosed with prostate carcinoma : an assessment of the validity and reliability of a self-administered questionnaire developed to measure the needs experienced by men diagnosed with prostate carcinom. *Cancer* 97, 993–1001. doi: 10.1002/cncr.11156

Fabrigar, L. R., Wegener, D. T., Maccallum, R. C., and Strahan, E. J. (1999). Evaluating the Use of exploratory factor analysis in psychological research. *Psychol. Methods* 4:272. doi: 10.1037/1082-989x.4.3.272

Guo, R., Xie, D. Q., Qi, Y., Zhang, R., and Zhang, J. (2019). Effects of personalized psychological intervention on adverse mood and quality of life in patients with lung cancer. *Chin. J. Health Psychol.* 27, 577–580.

Hilden, J. M., Emanuel, E. J., Fairclough, D. L., Link, M. P., Foley, K. M., Clarridge, B. C., et al. (2001). Attitudes and practices among pediatric oncologists regarding end-of-life care: results of the 1998 American Society of Clinical Oncology survey. *J. Clin. Oncol.* 19, 205–212. doi: 10.1200/jco.2001.19.1.205

Hodgkinson, K., Butow, P., Hunt, G. E., Pendlebury, S., Hobbs, K. M., Lo, S. K., et al. (2007). The development and evaluation of a measure to assess cancer survivors' unmet supportive care needs: the CaSUN (Cancer Survivors' Unmet Needs measure). *Psycho Oncol.* 16, 796–804. doi: 10.1002/pon.1137

Johnsen, A. T., Petersen, M. A., Pedersen, L., and Groenvold, M. (2011). Development and initial validation of the three-levels-of-needs questionnaire for self-assessment of palliative needs in patients with cancer. *J. Pain Symp. Manag.* 41, 1025–1039. doi: 10.1016/j.jpainsymman.2010.08.013

Kangas, M., Henry, J. L., and Bryant, R. A. (2002). Posttraumatic stress disorder following cancer. A conceptual and empirical review. *Clin. Psychol. Rev.* 22, 499–524. doi: 10.1016/s0272-7358(01)00118-0

Koltko-Rivera, M. E. (2006). Rediscovering the later version of Maslow's hierarchy of needs: Self-transcendence and opportunities for theory, research, and unification. *Rev. Gen. Psychol.* 10, 302–317. doi: 10.1037/1089-2680.10.4.302

Kottner, J., and Streiner, D. L. (2010). Internal consistency and Cronbach's α: a comment on Beeckman et al. (2010). *Int. J. Nurs. Stud*. 47, 926–928.

Li, W. W. Y., Lam, W. W. T., Shun, S. C., Lai, Y. H., Law, W. L., Poon, T., et al. (2013). Psychometric assessment of the chinese version of the supportive care needs survey short-form (SCNS-SF34-C) among Hong Kong and Taiwanese Chinese Colorectal Cancer Patients. *PLoS One* 8:e75755. doi: 10.1371/journal.pone.0075755

Lloret, S., Ferreres, A., Hernández, A., and Tomás, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anal. Psicol.* 30, 1151–1169.

Marijnen, C., van de Velde, C., Putter, H., van den Brink, M., Maas, C. P., Martijn, H., et al. (2005). Impact of short-term preoperative radiotherapy on health-related quality of life and sexual functioning in primary rectal cancer: report of a multicenter randomized trial. *J. Clin. Oncol.* 23, 1847–1858. doi: 10.1200/JCO.2005.05.256

Maslow, A. H. (1943). A theory of human motivation. *Psychol. Rev.* 50, 370–396. doi: 10.1037/h0054346

Massie, M. J. (2004). Prevalence of depression in patients with cancer. *J. Natl. Cancer Inst. Monogr.* 2004, 57–71.

Muthén, L. K., and Muthén, B. (2013). *Mplus User's Guide*, 7 Edn. Los Angeles, CA: Muthén and Muthén.

Peng, Y., Huang, M., and Kao, C. (2019). Prevalence of depression and anxiety in colorectal cancer patients: a literature review. *Int. J. Environ. Res. Public Health* 16:411. doi: 10.3390/ijerph16030411

Pigott, C., Pollard, A., Thomson, K., and Aranda, S. (2009). Unmet needs in cancer patients: development of a supportive needs screening tool (SNST). *Support. Care Cancer* 17, 33–45. doi: 10.1007/s00520-008-0448-7

Pirl, W. F. (2004). Evidence report on the occurrence, assessment, and treatment of depression in cancer patients. *J. Natl. Cancer Inst. Monogr.* 32, 32–39. doi: 10.1093/jncimonographs/lgh026

Price, M. A., Tennant, C. C., Butow, P. N., Smith, R. C., and Dunn, S. M. (2001). The role of psychosocial factors in the development of breast carcinoma: Part II. Life event stressors, social support, defense style, and emotional control and their interactions. *Cancer* 91, 686–697. doi: 10.1002/1097-0142(20010215)91:4<686::aid-cncr1052>3.0.co;2-0

Richards, C. T., Gisondi, M. A., Chang, C. H., Courtney, D. M., Engel, K. G., Emanuel, L., et al. (2011). Palliative care symptom assessment for patients with cancer in the emergency department: validation of the screen for palliative and end-of-life care needs in the emergency department instrument. *J. Palliat. Med.* 14, 757–764. doi: 10.1089/jpm.2010.0456

Richardson, A., Medina, J., Brown, V., and Sitzia, J. (2007). Patients' needs assessment in cancer care: a review of assessment tools. *Support. Care Cancer* 15, 1125–1144. doi: 10.1007/s00520-006-0205-8

Ryan, R. M., and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* 55, 68–78. doi: 10.1037/0003-066x.55.1.68

Sollner, W., Maislinger, S., Konig, A., Devries, A., and Lukas, P. (2004). Providing psychosocial support for breast cancer patients based on screening for distress within a Consultation-Liaison service. *Psycho Oncol.* 13, 893–897. doi: 10.1002/pon.897

Stanton, A. L., Bower, J. E., and Low, C. A. (2006). "Posttraumatic growth after cancer," in *Handbook of Posttraumatic Growth: Research and Practice*, eds L. G. Calhoun and R. G. Tedeschi (Hillsdale, NJ: Lawrence Erlbaum Associates), 138–175, (Reprinted).

Stark, D. P. H., and House, A. (2000). Anxiety in cancer patients. *Br. J. Cancer* 83, 1261–1267. doi: 10.1054/bjoc.2000.1405

Strong, V., Waters, R., Hibberd, C., Murray, G., Wall, L., Walker, J., et al. (2008). Management of depression for people with cancer (SMaRT oncology 1): a randomised trial. *Lancet* 372, 40–48. doi: 10.1016/S0140-6736(08)60991-5

Taber, K. S. (2017). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res. Sci. Educ.* 1, 1–24.

Tavakol, M., and Dennick, R. (2011). Making sense of Cronbach's alpha. *Int. J. Med. Educ.* 2, 53–55.

Walker, M. S., Ristvedt, S. L., and Haughey, B. H. (2003). Patient care in multidisciplinary cancer clinics: does attention to psychosocial needs predict patient satisfaction? *Psycho Oncol.* 12, 291–300. doi: 10.1002/pon.651

Wei, M., Shaffer, P. A., Young, S. K., and Zakalik, R. A. (2005). Adult attachment, shame, depression, and loneliness: the mediation role of basic psychological needs satisfaction. *J. Counsel. Psychol.* 52, 591–601. doi: 10.1037/0022-0167.52.4.591

WHO (2020). *World Health Statistics 2020: Monitoring Health for the SDGs, Sustainable Development Goals*. Geneva, (Reprinted).

WHO (2021). *Palliative Care*. Available online at: https://www.who.int/cancer/palliative/zh/# (accessed January 1, 2021).

Wu, J. L., and Xiao, J. X. (2011). The mechanism of cancer related depression. *Mod. Oncol. Med.* 19, 380–382.

Yogaparan, T., Panju, A., Minden, M., Brandwein, J., Mohamedali, H. Z., and Alibhai, S. (2009). Information needs of adult patients 50 or older with newly diagnosed acute myeloid leukemia. *Leuk. Res.* 33, 1288–1290. doi: 10.1016/j.leukres.2008.12.008

Yu, C., Li, X., Wang, S., and Zhang, W. (2016). Teacher autonomy support reduces adolescent anxiety and depression: an 18-month longitudinal study. *J. Adolesc.* 49, 115–123. doi: 10.1016/j.adolescence.2016.03.001

Zebrack, B. (2009). Information and service needs for young adult cancer survivors. *Support. Care Cancer* 17, 349–357. doi: 10.1007/s00520-008-0469-2

Zhao, X. J., Yang, Y., Wu, A. P., and Sheng, Y. (2015). Investigation and analysis of hospice care needs of patients with terminal cancer. *J. Nurs.* 30, 27–30.

Zung, W. (1965). A self-rating depression scale. *Arch. Gen. Psychiatry* 12:63.

Zung, W. (1971). Rating instrument for anxiety disorders. *Psychosomatics* 12, 371–379. doi: 10.1016/s0033-3182(71)71479-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 28 July 2021
doi: 10.3389/fpsyg.2021.663035

Psychometric Evaluation of the Chinese Recovering Quality of Life (ReQoL) Outcome Measure and Assessment of Health-Related Quality of Life During the COVID-19 Pandemic

Richard Huan Xu[1,2], *Anju Devianee Keetharuth*[3], *Ling-ling Wang*[4], *Annie Wai-ling Cheung*[2] and *Eliza Lai-yi Wong*[2]*

[1] Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, [2] Centre for Health Systems and Policy Research, Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, Hong Kong, [3] School of Health and Related Research, The University of Sheffield, Sheffield, United Kingdom, [4] Blood Transfusion Department, Jinling Hospital, School of Medicine, Nanjing University, Nanjing, China

OPEN ACCESS

Edited by:
Paola Gremigni,
University of Bologna, Italy

Reviewed by:
Agustín Simonelli Muñoz,
University of Almería, Spain
Ana Maria Hernandez Baeza,
University of Valencia, Spain

***Correspondence:**
Eliza Lai-yi Wong
lywong@cuhk.edu.hk

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: 02 February 2021
Accepted: 14 June 2021
Published: 28 July 2021

Citation:
Xu RH, Keetharuth AD, Wang L-l, Cheung AW-l and Wong EL-y (2021) Psychometric Evaluation of the Chinese Recovering Quality of Life (ReQoL) Outcome Measure and Assessment of Health-Related Quality of Life During the COVID-19 Pandemic. Front. Psychol. 12:663035. doi: 10.3389/fpsyg.2021.663035

Objective: The primary objective was to translate the Recovering Quality of Life (ReQoL) measures from English to traditional Chinese and assess their psychometric properties in Hong Kong (HK) Chinese population. The secondary objective was to investigate the mental health-related quality of life (HRQoL) of this sample during the coronavirus disease 2019 (COVID-19) pandemic.

Method: Recovering Quality of Life was translated to Traditional Chinese adhering to standard guideline recommended by the official distributors. Five hundred members of the general population were successfully recruited to participate in a telephone-based survey. The following psychometric properties of the ReQoL were evaluated: construct, convergent, and known-group validity and internal consistency and test–retest reliability. The item measurement invariance was assessed on the basis of differential item functioning (DIF). Multiple regression analysis was used to assess the relationship between respondents' characteristics and mental HRQoL.

Results: Results of confirmatory factor analysis (CFA) supported a two-factor structure of the ReQoL. The ReQoL showed significant correlations with the other mental health, quality of life, and well-being measures, which indicated a satisfactory convergent validity. Known-group validity confirmed that ReQoL is able to differentiate between people with different mental health status. The (Cronbach's alpha = 0.91 and 0.76 for positive [PF] and negative [NF] factor), and McDonald's omega of 0.89 (PF = 0.94, NF = 0.82) indicated the ReQoL has good reliability as well as test–retest reliability with an intraclass correlation coefficient of 0.75. Four items showed negligible DIF with respect to age. Respondents who were highly educated and without psychological problems reported a high ReQoL score.

Conclusion: Traditional Chinese ReQoL was shown to be a valid and reliable instrument to assess the recovery-focused quality of life in HK general population. Future studies are needed to appraise its psychometric properties in local people experiencing mental disorders.

Keywords: health-related quality of life, mental health, recovery, China, validation, ReQoL, psychometric properties

INTRODUCTION

Health-related quality of life (HRQoL) captures an individual's or population's perceived physical and mental health status over time (Wong et al., 2020). It can provide comprehensive information on the burden of preventable diseases, injuries, and disabilities from the perspective of person-centered care (Yin et al., 2016). Usually, HRQoL is assessed by using patient reported-outcome measures (PROMs), which include multiple items reflecting people's self-perceived physical and emotional functioning and health status. HRQoL has more traditionally been an important outcome to assess the effectiveness of interventions on people' physical health (Goldhagen et al., 2016; Xu et al., 2017; Wong E.L.Y. et al., 2019). Recently, however, in evaluating the outcomes of mental care, promoting "recovery," which reflects the extent to predict the changes in HRQoL (Garner et al., 2014), has drawn increasing professional attention, and emerged as a new paradigm to assess the full journey of people in overcoming the detrimental effects of mental problems (Ellison et al., 2016). Mental health recovery is a self-directed process of healing and transformation (Deegan, 2002), which is promoted through an interaction between individual experience, community environment and social engagement.

Although several PROMs are available to assess the effectiveness of interventions that reduce the symptoms of mental illness, few of them focus on appraising the improvement of mental HRQoL (Keetharuth et al., 2018a). Currently, the EuroQol-five dimension instrument (EQ-5D), a generic preference-based measure, is the most recommended PROM in assessing people's HRQoL worldwide (Herdman et al., 2011) and has been increasingly used to measure HRQoL in different populations in HK. However, the EQ-5D focused on physical health with only one of five items capturing mental health. In mental health, validity of the EQ-5D is rarely reported but suggests a potentially mixed picture (Brazier, 2010). Although there is evidence that generic instruments are able to reflect the impact of common conditions such as mild to moderate depression and anxiety (Lamers et al., 2006), an increasing number of studies showed conflicting evidence on its validity for patients with schizophrenia, depression, and bipolar (Barton et al., 2009; Papaioannou et al., 2011; Mulhern et al., 2014). Therefore, the Recovering Quality of Life (ReQoL) outcome measure with an emphasis on mental health was constructed to assess the recovery-focused quality of life (Keetharuth et al., 2018b).

Recovering Quality of Life was developed by a team led from The University of Sheffield, United Kingdom and funded by the Department of Health Policy Research Program (ReQoL, 2021). It is a self-completed questionnaire with two versions, ReQoL-20 and ReQoL-10, which contains 20 and 10 mental health items and one physical health item (not included in the scoring), respectively. The ReQoL was developed with significant inputs from service users not only as participants but also as research partners. The psychometric analyses in the development stage were based on data from over 6,450 participants, which significantly increased the face and content validity of the measure (Keetharuth et al., 2018b). A bifactor model of the ReQoL comprising a global factor and two local factors of negative and positive affects was reported by the developers (Keetharuth et al., 2018a). ReQoL has been translated into different languages and shown good reliability and validity (Keetharuth et al., 2018b; Chua et al., 2020; van Aken et al., 2020).

At present, widespread outbreaks of coronavirus disease 2019 (COVID-19) have drastically changed multiple aspects of the people's lives, with a growing number struggling with several mental health issues (Khan et al., 2020). Mass lockdown, increased rate of unemployment and the fears and uncertainties of the pandemic have not only exacerbated the psychiatric symptoms for patients with mental illnesses, but also affected the general population who may not have previously experienced psychological distress and symptoms of mental illness (Shigemura et al., 2020; Zhang and Ma, 2020). Chew et al. (2020) indicated that these psychological responses affect the well-being of the individual and community, and the impact could persist well after the outbreak. Traditional mental health measures, e.g., General Anxiety Disorder-7 (GAD-7) and Patient Health Questionnaire (PHQ-9), were developed on the basis of meeting clinical criteria and assess the effectiveness of intervention from the perspective of reducing symptoms (Keetharuth et al., 2018b). They cannot be used to reflect the lived experience of general population and measure their HRQoL outcomes. The lack of a well-defined, multidimensional, and psychometrically valid measure of recovery in Hong Kong (HK) has been cited as a potential barrier to providing suitable healthcare for improving their mental HRQoL (Mak et al., 2016). Although the ReQoL was constructed to assess the HRQoL for patients with a broad spectrum of mental illnesses, it could be a useful instrument to capture general wellbeing in the pandemic as well as establishing whether this questionnaire could be used for public health interventions in the general population. Therefore, the primary objective of this study was to translate and culturally adapt the ReQoL from English to Traditional Chinese (ReQoL-TC) and assess its psychometric properties in HK general population. The secondary aim was to investigate the recovery-focused HRQOL of this sample using the ReQoL-TC during the COVID-19 pandemic.

MATERIALS AND METHODS

Translation and Cultural Adaptation

We adhered to standard guidelines "Translation and Linguistic Validation Process for the ReQoL" provided by the official distributors of the ReQoL in translating and culturally adapting the ReQoL in Traditional Chinese (Wild et al., 2005). Dual forward translation was undertaken independently by two professional translators, who were native Chinese speakers but proficient in English. The local research team used both translations to perform the forward translation reconciliation. A revised version was then produced and sent to another two professional translators, who were native English speakers but proficient in Chinese for backward translation independently. The local and ReQoL research team jointly examined the back translation against the original English version to identify any discrepancies, addressed the disputed items, and refined the translation focused on cultural adaption until consensus was achieved by all the research team members.

Cognitive debriefing was conducted with ten members of the general population who were invited to comment on the response options and any wording they found difficult in understanding the ReQoL-TC. Respondents were asked to describe in their own language what the wording meant to them. We paid special attention to the items that we had to adapt and were confident they were understood in the way intended as per the developer's concept elaboration. After proofreading, the final version of the ReQoL-TC was confirmed.

The local research team has rich experience in the translation and cultural adaption of PROMs and other health outcome-related questionnaires. The eligibility of the translators has been approved by the ReQoL distributor. The results were discussed with the ReQoL developers and one of whom was invited to join local research team to monitor the project and ensure the quality of the development.

Sample and Data Collection

A random telephone survey of the general population in HK, was carried out by a team of telephone survey professionals in July 2020, following similar recruitment methods by a previous study (Chan et al., 2019). Before the formal survey, a pilot study with ten randomly selected persons was conducted to test the logic of the telephone survey. In order to minimize the sampling error, telephone numbers were first selected randomly from an updated telephone directory as seed numbers. Another three set of numbers were then generated using the randomization of last two digits in order to recruit the unlisted numbers. Duplicate numbers were screened out, and the remaining numbers mixed in a random order to form the final sample. Interviews were carried out by experienced interviewers, between 10:00 and 22:00 on weekdays and other periods including weekends and public holidays should appointments with suitable subjects were arranged. The inclusion criteria for the study were: (1) HK permanent residents; (2) $\geq$18 years; (3) no cognitive problems; and (4) able to provide informed consent. Upon successful contact with a target household, one qualified member of the household was selected among family members using the last-birthday random selection method (i.e., a respondent in the household who just had their birthday would be selected to participate in the telephone interview). Given a sample size of around 300–500 is believed to have sufficient power to estimate parameters in confirmatory factor analysis (CFA; DeVellis, 2017), in this study, a minimum sample size of 500 was determined. Finally, data from 500 participants were successfully collected. Approximately 72.2% were female, around 60.6% $\geq$60 years, nearly 64.4% completed the secondary or above education, 62.2% of participants reported no chronic conditions and only 4.4% (n = 22) indicated they had visited a psychiatrist within the last 12 months (**Table 1**). The flowchart of the participant recruitment and selection process is presented in **Supplementary Figure 1**. Study protocol and informed consent was approved by the institutional review board of The Chinese University of Hong Kong (Ref. ID: SBRE-18-671).

Measurement

Recovering Quality of Life

The ReQoL-TC was used in this study. It comprises 20 mental health items and one physical health item. ReQoL-TC-10 (11 items) comprises the first 10 item of the ReQoL-TC and the physical health item. Of the 20-item ReQoL-TC, 11 are positively worded and nine are negatively worded. All the items are scored on a five-point Likert scale ranging from "None of the time" to "Most or all of the time." A sum score is calculated by summing the scores of all the items (except for the physical health item), where a higher score indicates a better quality of life.

General Anxiety Disorder-7

The GAD-7 is a self-rated scale to measure the severity of generalized anxiety disorder. It has seven items, e.g., Feeling nervous, anxious, or on edge, scored from zero (not at all) to three (nearly every day) (Kroenke et al., 2007). The sum score ranges from zero to 21 and the cut-off point for mild, moderate and severe anxiety symptoms are 5, 10, and 15, respectively (Spitzer et al., 2006). The Chinese GAD-7 has been validated (Tong et al., 2015). The Cronbach's alpha (internal consistency reliability) of the GAD-7 in our sample was 0.93.

Depression Anxiety Stress Scales-21 Items

The Depression Anxiety Stress Scales-21 items (DASS-21) measures higher-order mental factor of psychological distress (Lovibond and Lovibond, 1995) over the past week using seven items in each of the domains of depression, anxiety, and stress. Each item, e.g., I found it hard to wind down, has a four-point Likert scale with rating choices ranging between "never applied to oneself" (0) and "very much/most of the time" (3). Final scores for three subscales are calculated by summing the scores for the relevant items (DASS, 2021). The Chinese DASS-21 has been validated (Wang et al., 2016). The Cronbach's alpha (internal

TABLE 1 | Participants' characteristics.

	n	%
Overall	500	100
Sex		
Male	139	27.8
Female	361	72.2
Age		
18–49	107	21.4
50–59	90	18.0
60–69	138	27.6
≥70	165	33.0
Educational level		
Primary or below	174	34.8
Secondary/post-secondary	216	43.2
Tertiary or above	106	21.2
Marital status		
Single	67	13.4
Married	381	76.2
Divorced/widow(er)	50	10.0
Living status		
Living alone	52	10.4
Living with families	448	89.6
Working status		
Fully employed	139	27.8
Non-employed	179	35.8
Retired	182	36.4
Government allowance		
Receiver	177	35.4
Non-receiver	323	64.6
Personal income per month (HK dollar) [1HKD = 0.13 US dollar]		
≤5,000	313	62.6
5,001–20,000	93	18.6
≥20,001	45	9.0
Refused	49	9.8
Chronic conditions		
Yes	189	37.8
No	311	62.2
Mental health problem		
Yes	22	4.4
No	478	95.6

Non-employed includes respondents who reported they are housewife, students, and unemployed.

consistency reliability) of subscale depression, anxiety and stress in our sample was 0.75, 0.7, and 0.72, respectively.

EQ-5D-5L

The EQ-5D-5L is a generic preference-based measure to estimate people's HRQoL (Herdman et al., 2011). It has two sections: the descriptive system and the visual analog scale (EQ-VAS). The descriptive system comprises five items (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with five levels (from "no problem" to "extreme problem"). Utility scores were calculated using the EQ-5D-5L HK value set (Wong et al., 2018; Wong E.L. et al., 2019). EQ-VAS is a vertical scale used to measure people's overall health with values between 0 (worst imaginable health) and 100 (best imaginable health) The Cronbach's alpha (internal consistency reliability) of the EQ-5D-5L (descriptive system) in our sample was 0.82.

ICEpop CAPability Measure for Adults

ICEpop CAPability measure for adults (ICECAP-A) is a generic preference-based measure that evaluates an individual's capability well-being (Al-Janabi et al., 2012). The descriptive system of the ICECAP-A has five items (stability, attachment, autonomy, achievement, and enjoyment) with four response options ranging from "fully capable" to "not capable." The Chinese ICECAP-A has been validated (Tang et al., 2018). In the absence of value set for the Chinese population, we calculated the sum score of the ICECAP-A by summing the scores of five items, where a higher score represents a poorer capability well-being. The Cronbach's alpha (internal consistency reliability) of the ICECAP-A in our sample was 0.77.

Statistical Analysis

Construct, Convergent, and Known-Group Validity

Confirmatory factor analysis was used to assess the construct validity. In assessing the dimensionality of the ReQoL-TC, two models were developed in line with the original study (Keetharuth et al., 2018a). The first was a bi-factor model, with one global factor and the two factors contained positively worded and negatively worded items respectively. Second, the two-factor model consisted of the positively worded items and the negatively worded items as two separate factors. The model fit was assessed by the root mean square error of approximation (RMSEA $\leq$ 0.08), standardized root mean squared residual (SRMR $\leq$ 0.08), Tucker-Lewis index (TLI $\geq$ 0.9), and comparative fit index (CFI $\geq$ 0.9) (DeVellis, 2017). The factor loadings of each item were also checked. For the bi-factor model, we calculated the explained common variance for the global factor to assess its importance relative to the two other factors (Reise et al., 2010). To address the issue of non-normal data, the robust distribution free weighted least squares (WLSMV) estimator was used (Ferrando and Lorenzo-Seva, 2000; Markon, 2019).

Several *a priori* hypotheses about the relationship between the ReQoL-TC and the other HRQoL and mental health instruments were formulated to test the convergent validity of the ReQoL-TC (Keetharuth et al., 2018b; Chua et al., 2020; van Aken et al., 2020) (specific hypotheses are presented in **Supplementary Table 1**). The strength of the correlation was estimated using Pearson's correlation coefficient (r), where $r \geq 0.55$ were interpreted as adequate.

The known-group validity was examined by (i) comparing people self-reporting specific mental health conditions versus those who did not; and (ii) using GAD-7 and DASS clinical cut-off points (where a score of <5 on GAD-7 (Spitzer et al., 2006) and depression [≤9], anxiety [≤7], and stress [≤14] on DASS indicate no clinical concerns) (Brumby et al., 2011). While GAD-7 and DASS-21 do not measure aspects of quality of life *per se*, it can be assumed that they define broad groups expected to have different quality of life scores.

TABLE 2 | Fit statistics from confirmatory factor analytic models.

	Bi-factor model: global, negative, positive	Two-factor model: negative and positive
χ^2	214.7	540.3
Degree of freedom	150	169
p-value	<0.001	<0.001
RMSEA (<0.08)	0.029	0.066
SRMR (<0.08)	0.058	0.076
TLI (>0.9)	0.989	0.942
CFI (>0.9)	0.991	0.948

Item Statistics, Internal Consistency, and Test–Retest Reliability

Feasibility and acceptability of the ReQoL-TC were assessed by the time taken to complete the questionnaire and proportions of missing values of items (respondents were allowed to skip questions) (Xu et al., 2018). The mean, standard deviation (SD), median, and range of the ReQoL-TC (both 20- and 10-item versions) scores were reported. We also calculated ceiling and floor effects, skewness, and kurtosis. Internal consistency reliability of the ReQoL-TC was measured by Cronbach's alpha ($\alpha > 0.7$, acceptable), McDonald's omega ($\omega > 0.7$, acceptable), Guttman's lambda 4 ($\lambda > 0.8$, suitable) (McDonald, 1999), the item-total correlation (>0.5, acceptable) and alpha if an item is dropped (DeVellis, 2017). Selected participants (10%) was invited to take part in another telephone survey two weeks later (only respondents who did not report experiencing any significant life event were invited) to assess the test–retest reliability using intraclass correlation coefficient (ICC > 0.7, acceptable, two-way mixed effects model) (Fleiss, 1999).

Differential Item Functioning

Differential item functioning (DIF) with regard to patients' natural attributes, i.e., gender (male vs. female) and age (regrouped to two groups; G1: < 50 years vs. G2: ≥ 50 years), which were unchanged characteristics, was evaluated (Cherepanov et al., 2010; Roberts et al., 2014). Three ordinal logistic regression models (Model 1: explanatory variable; Model 2: explanatory variable plus vector of group identifiers; and Model 3: explanatory variable multiplied by vector of group identifiers) were developed. The likelihood ratio [χ^2] was used test to compare the nested models (Robinson et al., 2019). A significant difference (Bonferroni correction was used) between Model 1 and Model 2, and Model 2 and Model 3 indicates the presence of uniform and non-uniform DIF, respectively (Choi et al., 2011). The degree of DIF was assessed on the differences in McFadden's pseudo-R^2 values (with corresponding effect sizes < 0.13, negligible; 0.13–0.26, moderate; and >0.26, large) (Zumbo, 1999). DIF analyses were conducted for the two dimensions differentiated in the ReQoL-TC.

The univariate (one-way analysis of variance) and multiple analysis (ordinary least squares regression model) was used to show the ReQoL-TC sum scores reported by respondents with different background characteristics, i.e., sex, age, education,

TABLE 3 | Results of CFA for the ReQoL-TC.

Item description	Item no	Bifactor model standardized loadings			Two-factor model standardized loadings	
		Global	NF	PF	NF	PF
I found it difficult to get started with everyday tasks	r1	0.19	0.39		0.49	
I felt able to trust others	r2	0.34		0.50		0.53
I felt unable to cope	r3	0.26	0.34		0.51	
I could do the things I wanted to do	r4	0.73		0.37		0.82
I felt happy	r5	0.90		0.04		0.74
I thought my life was not worth living	r6	0.09	0.30		0.34	
I enjoyed what I did	r7	0.84		0.25		0.85
I felt hopeful about my future	r8	0.72		0.33		0.80
I felt lonely	r9	0.33		0.32	0.38	
I felt confident in myself	r10	0.66		0.55		0.84
I did things I found rewarding	r11	0.72		0.47		0.86
I avoided things I needed to do	r12	0.01	0.37		0.36	
I felt irritated	r13	0.13	0.61		0.60	
I felt like a failure	r14	0.11	0.30		0.29	
I felt in control of my life	r15	0.64		0.55		0.82
I felt terrified	r16	0.23	0.64		0.58	
I felt anxious	r17	0.01	0.63		0.65	
I had problems with my sleep	r18	0.03	0.47		0.43	
I felt calm	r19	0.20	0.00	0.58		0.42
I found it hard to concentrate	r20	0.02	0.59		0.37	

NF – factor made up of negatively worded items; PF – factor made up of positively worded items.

marital status, living status, working status, chronic disease status, and mental health condition, respectively. R software was used for all the statistical analyses (R Core Team, 2013). CFA, reliability, ICC and DIF was analyzed using "lavaan," "psych," "ICC," and "lordif" package, respectively. The level of significance was set at p-value ≤ 0.05.

RESULTS

The model fit statistics from the CFA are presented in **Table 2**. The goodness-of-fit indices indicated an acceptable fit for both the bifactor ($\chi^2 = 214.7$, degree of freedom [df] = 150, $p < 0.001$, RMSEA = 0.029, SRMR = 0.058, TLI = 0.989, and CFI = 0.991) and the two-factor model ($\chi^2 = 540.3$, df = 169, $p < 0.001$, RMSEA = 0.066, SRMR = 0.076, TLI = 0.942, and CFI = 0.948) of the ReQoL-TC. The factor loadings for the two-factor model ranged between 0.29 and 0.86. For the bi-factor model, the loadings for the global factor were smaller than 0.3 for 11 out of 20 items (40 factor loadings: 10 for positive factor, 10 for negative factor, and 20 for global factor) and the explained common variance was 51%. Considering the results of CFA and the bi-factor structure reported by the original study, we concluded that the bi-factor model outperformed the two-factor model in this sample of HK general population. The standardized factor loadings for the observed variables of both models are presented in **Table 3**.

The distribution of the ReQoL-TC sum and factor scores are presented in **Table 4**. The mean scores (SD) for the ReQoL-TC-20 and the ReQoL-TC-10 were 60.33 (10.52) and 28.54 (6.63), respectively, and no sum score of three scales showed ceiling or floor effect. The internal consistency reliability of ReQoL-TC-20 (α = 0.86 [0.91 and 0.76 for two factors, respectively]) and ReQoL-TC-10 (α = 0.84) was acceptable. The value of ICC confirmed the test–retest reliability of ReQoL-TC-20 ($ICC_{overall} = 0.75$, $ICC_{positive} = 0.79$, and $ICC_{negative} = 0.71$) and ReQoL-TC-10 (ICC = 0.71) was satisfactory. No missing data was identified of ReQoL-TC and the average time to complete the measure was around 5 min indicating a good feasibility and acceptability. The response distribution, and factor-level item-total correlation and alpha if item dropped of the ReQoL-TC are presented in **Supplementary Table 3**. The sum and factor score distributions of the ReQoL-TC are presented in **Supplementary Figure 2**.

Tables 5, **6** show the result of convergent and known-group validity of the ReQoL-TC. All the correlations between measures were significant and the signs were as expected. Both the ReQoL-TC-20 and ReQoL-TC-10 showed a significant correlation with GAD-7, DASS-21, EQ-5D, and ICECAP-A scores. Most correlation coefficients of the ReQoL-TC-20 were larger than those of the ReQoL-TC-10. The ICECAP-A item of enjoyment ($r = -0.49$) showed the strongest correlation with the sum score of ReQoL-TC-20, followed by the ICECAP-A item of stability ($r = -0.47$) and attachment ($r = -0.47$). The correlation coefficients of the ReQoL-TC-10 with the other measures ranged between 0.23 (GAD-7) and 0.48 (ICECAP-A item of attachment). The results of ANOVA indicated that participants with clinical mental health status showed poorer quality of life than those without, confirming the known-group validity of the ReQoL-TC.

Factor-level DIF analysis found that items 14 (negative factor) showed both uniform and non-uniform DIF on sex. Another three negatively worded items 6, 10 (uniform), and 9 (non-uniform) showed DIF on age. Two positively worded items 3 (uniform) and 8 (non-uniform) showed DIF on age. Checking the McFadden R^2, the effect size of DIF was negligible for all six items (<0.001–0.06) (**Supplementary Table 4**).

Results of the univariate and multiple analysis are presented in **Table 7**. Respondents who were highly educated and living with no psychological problems tended to report a high ReQoL-TC sum score.

DISCUSSION

This study presented the development of the Traditional Chinese version of the ReQoL, which exhibited acceptable psychometric properties, in HK general population. The translation process adhered to acceptable international translation standard. A two-factor structure, which separately comprised positively and

TABLE 4 | Score distribution and internal consistency and test–retest reliability of ReQoL-TC.

	ReQoL-20 (scale 0–80)	ReQoL-10 (scale 0–40)	ReQoL (physical item)
Measure statistics			
Mean	60.33	28.54	3.57
Standard deviation	10.52	6.63	0.71
Median	61	30	4
Range	22–80	5–40	0–4
Ceiling effect %	0.2	0.6	68.0
Floor effect %	0.2	0.2	0.4
Skewness	−0.42	−0.44	−1.78
Kurtosis	−0.33	−0.58	3.3
Reliability			
Overall			
Cronbach's alpha	0.86	0.84	
McDonald's omega	0.89	0.86	
Guttman's lambda 4	0.92	0.88	
ICC (95% C.I.)	0.75 (0.6–0.84)	0.71 (0.54–0.82)	
Positive domain			
Cronbach's alpha	0.91		
McDonald's omega	0.94		
Guttman's lambda 4	0.94		
ICC (95% C.I.)	0.79 (0.68–0.87)		
Negative domain			
Cronbach's alpha	0.76		
McDonald's omega	0.82		
Guttman's lambda 4	0.86		
ICC (95% C.I.)	0.71 (0.62–0.81)		

ICC, intraclass correlation coefficient.

TABLE 5 | Convergent validity of the ReQoL-TC.

	ReQoL-20			ReQoL-10
	Overall	**Positive**	**Negative**	
GAD-7	−0.36***	−0.2***	−0.52***	−0.23***
DASS-21				
Depression	−0.44***	−0.21***	−0.62***	−0.33***
Anxiety	−0.38***	−0.16***	−0.56***	−0.26***
Stress	−0.43***	−0.23***	−0.58***	−0.34***
EQ-5D				
Anxiety/Depression item	−0.32***	−0.22***	−0.31***	−0.27***
Utility score	0.3***	0.26***	0.2***	0.3***
EQ-VAS	0.39***	0.3	0.32	0.34***
ICECAP-A				
Stability	−0.47***	−0.46***	−0.2***	−0.45***
Attachment	−0.47***	−0.46***	−0.19***	−0.48***
Enjoyment	−0.49***	−0.44***	−0.29***	−0.47***
Achievement	−0.35***	−0.3***	−0.24***	−0.35***
Autonomy	−0.26***	−0.29***	0.06	−0.26***
ReQoL-20				
Overall	-	-	-	0.95***
Positive	-	-	0.2***	–

*** $p < 0.001$.

negatively worded items, was confirmed with an acceptable model fit and factor loadings. We have not found commonly agreed threshold for interpreting the explained common variance. However, previous studies have concluded that scales were sufficiently unidimensional if they obtained ECV values of around 70–80% which is much higher than 51% found in this study (Reise et al., 2013). The ReQoL-TC showed good convergent validity in correlated with other instruments measuring HRQoL, mental health and well-being, and sufficient discriminative power to differentiate people with and without clinical mental health status as defined by GAD-7 and DASS-21. Additionally, the internal consistency and test–retest reliability of the ReQoL-TC were satisfactory for both positive wording and negative wording factors, and no ceiling or floor effect was detected. Further, a significant and strong correlation between ReQoL-TC-20 and ReQoL-TC-10 was identified. In general, the results of this study confirmed that two-factor ReQoL-TC is a valid and reliable instrument with good acceptability and feasibility to assess the recovery-focused quality of life for HK general population and can be used in research settings.

Although no ceiling or floor effects were detected on the sum score of ReQoL-TC, score distribution of some negatively worded items were severely skewed. Approximately 92 and 89% of participants chose the option "Never" for item "*I felt like a failure*" and "*I thought my life was not* worth living." This is in line with the findings from the original United Kingdom study where 33 and 51% of patients indicated never having any concerns on those two aspects. Additionally, item "*I felt in control of my life*" and item "*I felt calm*" showed a measure of floor effect with 31 and 25% of participants indicating they could control their life or feel calm, respectively, which was not reported in the original study. These findings should be interpreted with caution as our survey was completed during the COVID-19 pandemic. Given several studies have confirmed that pandemic undoubtedly negatively

TABLE 6 | Known-group validity of the ReQoL-TC.

	***N* = 500**	Mean (standard deviation)			
		ReQoL-20 (scale 0–80)	**ReQoL-20 negative**	**ReQoL-20 positive**	**ReQoL-10 (scale 0–40)**
Self-reported mental illness					
Yes	22	53.09	13.52(3.75)	25.32(8.44)	24.95
No	478	60.66	16.32(5.87)	29.66(8.79)	28.71
p-value		<0.001	<0.001	0.02	0.005
GAD-7					
No (<5)	437	61.69	20.22(5.2)	30(8.57)	29.13
Mild (5–10)	45	51.49	17.91(4.92)	26.22(8.7)	24.42
Moderate or above (≥10)	18	49.29	12.93(3.13)	25.64(10.3)	24.5
p-value		<0.001	<0.001	0.002	<0.001
DASS-depression					
Non-clinical (≤9)	472	61.16	21.79(4.88)	29.76(8.8)	28.92
Clinical (> 9)	28	46.36	13.16(3.26)	24.64(7.59)	22.18
p-value		< 0.001	<0.001	0.003	< 0.001
DASS-anxiety					
Non-clinical (≤7)	470	61.0	20.4(5.35)	29.64(8.87)	28.81
Clinical (>7)	30	49.77	13.21(3.37)	26.8(7.4)	24.4
p-value		<0.001	<0.001	0.08	<0.001
DASS-stress					
Non-clinical (≤14)	494	60.59	23.83(3.66)	25.59(5.39)	28.66
Clinical (>14)	6	39.17	13.52(3.74)	19.67(8.78)	18.5
p-value		<0.001	<0.001	<0.001	<0.001

impact on people's mental health and well-being (Burgess, 2020; Galea et al., 2020), the "COVID bias" might have affected the people's response in our study. Post-pandemic assessments are needed in the future.

The ReQoL-TC has a short version and a longer version to serve the different settings where recovery-focused quality of life is measured. Despite the internal consistency and test–retest reliability for both versions being satisfactory in this study, the Cronbach's alpha was lower than that reported by the studies in United Kingdom (ReQoL-20:0.93 and ReQoL-10:0.87) and Netherlands (ReQoL-20:0.94 and ReQoL-10:0.9) (Keetharuth et al., 2018b; van Aken et al., 2020). We found the mean sum score of the ReQoL-TC-10 was lower than that of the half of ReQoL-TC-20 sum score (scale 0–40). This difference might be because participants tend to score higher on those items with negative wording. ReQoL-TC-10 only contains four out of 11 items with negative wordings. However, previous studies indicated that a measure contains a mixture of positive and negative items is a crucial element as people with mental health difficulties identified issues that both enhanced or depleted their quality of life (Crawford et al., 2011; Keetharuth et al., 2018b). The psychometric performance of two versions of ReQoL-TC should be separately assessed in the future.

In our sample, compared with the ReQoL-TC-10, the ReQoL-TC-20 showed a closer relationship with the mental health-related measures, i.e., the GAD-7 and DASS-21. This is possibly the case because of the pandemic with more people experiencing mental health difficulties. This finding should be interpreted with caution because only 500 members of the general population were included in this study. However, overall, we could expect ReQoL-TC-10, by virtue of its brevity, to be more practical measuring the recovery-focused quality of life for general population. For individuals with mental illness, the ReQoL-TC-20 could be more appropriate due to its comprehensiveness. Future studies with both general population and individuals with mental problems are needed to investigate the generatability of findings in this study. Further, in line with the findings of van Aken et al.'s (2020) study, both ReQoL-20 and ReQoL-10 showed a significant but not strong correlation with the EQ-5D utility score. It is not surprising that, as Papaioannou et al. (2013) also indicated in their study, for patients with personality disorders, the EQ-5D is suitable to assess patients' HRQoL, but lacks the content validity to fully reflect the impact of the condition. Brazier also indicated that the EQ-5D appears to perform acceptably well in depression and personality disorder, but less well in anxiety, schizophrenia and bipolar disorder (Brazier et al., 2014). Furthermore, the ReQoL-TC sum score strongly correlated with ICECAP-A item of stability, attachment and enjoyment supported that ReQoL-TC can capture people' recovery-focused quality of life as it is intended to measure (Leamy et al., 2011), instead of assessing the effect of the reduction in symptoms alone.

The associations of positively and negatively worded factors of the ReQoL-TC with the other HRQoL and mental health measures showed a mixed picture. The items of negative wordings presented a strong correlation with the measures investigating individual's mental health status. However, the items of positive wordings factor showed a strong correlation with quality of life and well-being measures. Previous studies have indicated the potential impact of using negative or positive wordings in assessing individual's psychological attributes. For example, a study examined the Rosenberg Self-Esteem scale, has demonstrated the existence of method effects associated with negatively and/or positively worded items (Schönberger and Ponsford, 2009). Wouters et al. (2012) also indicated the

TABLE 7 | Univariable and multiple analysis of the ReQoL-TC.

	Mean (SD)	*p*-value	Coefficient (95% C.I.)
Overall	60.33 (10.52)		
Sex			
Male	61.12 (10.88)	0.3	Ref
Female	60.02 (10.38)		−2.04 (−4.29, 0.22)
Age			
18–49	60.9 (10.44)	0.6	Ref
50–59	59.92 (10.87)		0.06 (−3.46, 3.58)
60–69	59.48 (9.87)		2.62 (−1.07, 6.32)
≥70	60.89 (10.94)		7.22 (2.53, 11.91)**
Educational level			
Primary or below	59.07 (10.64)	<0.001	Ref
Secondary/post-secondary	60.27 (10.52)		2.55 (0.03, 5.08)*
Tertiary or above	63.04 (9.73)		4.09 (0.63, 7.55)*
Marital status			
Single	60.85 (8.64)	0.07	Ref
Married	60.66 (10.66)		−0.18 (−3.99, 3.62)
Divorced/widow (er)	57.08 (11.41)		−3.06 (−8.09, 1.97)
Living status			
Living alone	56.9 (10.48)	0.01	Ref
Living with families	60.73 (10.47)		1.59 (−2.11, 5.29)
Working status			
Fully-employed	60.57 (10.38)	<0.001	Ref
Non-employed	62.09 (9.68)		3.16 (−0.78, 7.1)
Retired	58.41 (11.15)		−3.07 (−7.48, 1.35)
Government allowance			
Receiver	59.74 (10.74)	0.4	Ref
Non-receiver	60.65 (10.4)		0.14 (−3.15, 3.43)
Personal income per month			
≤5,000	60.48 (10.52)	0.01	Ref
5,001–20,000	58.56 (10.41)		−1.01 (−4.84, 2.81)
≥20,001	63.93 (10.15)		3.43 (−1.34, 8.2)
Chronic conditions			
Yes	60.01 (10.58)	0.6	Ref
No	60.52 (10.5)		−0.05 (−2.46, 2.36)
Mental health problem			
Yes	53.09 (9.34)	<0.001	Ref
No	60.66 (10.46)		6.96 (2.23, 11.69)*

*Non-employed includes respondents who reported they are housewife, students, and unemployed. *p < 0.05; **p < 0.01.*

importance of evaluating wording effects when examining the factor structure of the HADS in vulnerable patient groups. However, few of them directly exhibited a significant association between items with positive wordings and respondents' well-being. Further research is needed to investigate the effect of negatively and positively worded items of the ReQoL-TC on the other Chinese population's health, and which socio-demographic and personality characteristics are associated with such response style.

Regarding the structure of the ReQoL-TC, Keetharuth et al.'s (2018b) original study reported a bi-factor model – a global factor and separate factors for the positively and negatively worded items. This finding was not fully supported in our study (low factor loadings for the global factor and a low explained common variance), despite the satisfactory goodness-of-fit indices. However, both two studies confirmed the presence of two factors – positively worded and negatively worded items – of the ReQoL. Moreover, our model fit was average as this might be as a result of the survey bias in terms of the interviewer administered questionnaire, the survey population and the influence of the COVID-19 pandemic. Considering the difference in the populations of the two studies (patient in the United Kingdom vs. general population in HK) and the goodness-of-fit of two models were satisfactory, we do not suggest rejecting the two-factor model. However, further assessments are needed to explore the structure of the ReQoL-TC in both HK general population or individuals experiencing mental illness. The implication of the two factors is that a sum score may not be generated for the ReQoL-TC for this population and the scores of positive and negative affect will have to be kept separate. Moreover, several items (four negative and one positive wording items) showed significant DIF on age and sex, respectively, despite the effect size of DIF was negligible.

Attention should be paid to the item "*I felt like a failure (item 14)*," which showed several psychometric problems, such as both uniform and non-uniform DIF, a low factor loading (0.29) and strong skewness (−5.32). This might be because of two possible reasons. First, "failure" is a very negative word in the Chinese culture and people usually show less willingness to use that word to describe themselves or the others (Bedford and Hwang, 2003), thus, in this study, more than 90% of respondents selected "none of the time". Second, the ReQoL-TC was developed based on inputs from individuals with mental health problems, however, in this study, all respondents were members of the general population, thus, negative wording items, such as "failure" item were problematic to some extent. However, considering this was the first study to assess the validity and reliability of the ReQoL-TC and only 500 members of the general population was surveyed, we retained all the items and did not recommend for any to be dropped at this stage.

Several limitations should be addressed. First, all participants in this study were recruited through a telephone-based survey. This might lead to several bias pertaining to data collection and quality, e.g., participants may not have understood the questions as they could not read them (interview bias). Hence, other forms, such as face-to-face or online survey, should be used in future studies. Second, our sample is not representative of the HK general population as few of them were young or with high income, which might affect the validity and reliability of the ReQoL-TC. Third, while we used the guidelines provided by the developers to adapt the ReQoL for the HK population, we are aware that there are newer guidelines on cultural adaptation (Hernandez et al., 2020). Moreover, despite the original ReQoL has been translated and adapted to HK Chinese population, the adaptation may not be directly used to compare between two cultures because no measurement equivalence between the original and the adapted forms was carried out. Further investigations are needed to independently assess the psychometric properties of the ReQoL-TC-10 in another sample of local population.

CONCLUSION

This study confirmed that the ReQoL-TC has sound psychometric properties in a sample of HK general population. It demonstrates good face and content validity, satisfactory convergent and discriminatory validity as well as adequate internal consistency and test–retest reliability. A bi-factor structure of the ReQoL-TC with one positive wording, one negative wording and a global factor was confirmed. This study also investigated the recovery-focused HRQoL using the newly developed ReQoL-TC in HK general population during the COVID-19 pandemic. Although we showed that ReQoL-TC is suitable for use in HK general population, future validation work should be carried out to investigate the performance of the ReQoL-TC in individuals with mental health problems. We also intend to develop a set of preference weights preference-based ReQoL-TC to calculate quality adjusted life years to support the economic evaluation in improving people's mental HRQoL.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by contacting correspondence author, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional review board of The Chinese University of Hong Kong (Ref. ID: SBRE-18-671). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RX: study concept and design, data analysis and interpretation, software, writing-original draft, and writing–review and editing. AK: study concept and design, data analysis and interpretation, and writing–review and editing. L-LW: software, visualization, and writing–review and editing. AW-LC: provision of study materials and patients, collection and assembly of data, and writing–review and editing. EL-YW: study concept and design, provision of study materials and patients, collection and assembly of data, supervision, and writing–review and editing. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.663035/full#supplementary-material

REFERENCES

Al-Janabi, H., Flynn, T. N., and Coast, J. (2012). Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Qual. Life Res.* 21, 167–176. doi: 10.1007/s11136-011-9927-2

Barton, G. R., Hodgekins, J., Mugford, M., Jones, P. B., Croudace, T., and Fowler, D. (2009). Measuring the benefits of treatment for psychosis: validity and responsiveness of the EQ–5D. *Br. J. Psychiatry* 195, 170–177.

Bedford, O., and Hwang, K.-K. (2003). Guilt and shame in chinese culture: a cross-cultural framework from the perspective of morality and identity. *J. Theory Soc. Behav.* 33, 127–144. doi: 10.1111/1468-5914.00210

Brazier, J. (2010). Is the EQ-5D fit for purpose in mental health? *Br. J. Psychiatry* 197:348. doi: 10.1192/bjp.bp.110.082453

Brazier, J., Connell, J., Papaioannou, D., Mukuria, C., Mulhern, B., Peasgood, T., et al. (2014). A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technol. Assess.* 18, 7–8. doi: 10.3310/hta18340

Brumby, S., Chandrasekara, A., McCoombe, S., Torres, S., Kremer, P., and Lewandowski, P. (2011). Reducing psychological distress and obesity in Australian farmers by promoting physical activity. *BMC Public Health* 11:362. doi: 10.1186/1471-2458-11-362

Burgess, R. (2020). COVID-19 mental-health responses neglect social realities. *Nature* doi: 10.1038/d41586-020-01313-9

Chan, C. W. H., Wong, M. H. H., Chan, K. W., Chow, A. Y. M., and Lo, R. S. Y. (2019). Prevalence, perception, and predictors of advance directives among Hong Kong Chinese: a population-based survey. *Int. J. Environ. Res. Public Health* 16:365. doi: 10.3390/ijerph16030365

Cherepanov, D., Palta, M., Fryback, D. G., and Robert, S. A. (2010). Gender differences in health-related quality-of-life are partly explained by Sociodemographic and socioeconomic variation between adult men and women in the US: evidence from four US nationally representative data sets. *Qual. Life Res.* 19, 1115–1125. doi: 10.1007/s11136-010-9673-x

Chew, Q. H., Wei, K. C., Vasoo, S., Chua, H. C., and Sim, K. (2020). Narrative synthesis of psychological and coping responses towards emerging infectious disease outbreaks in the general population: practical considerations for the COVID-19 pandemic. *Singapore Med. J.* 61, 350–356. doi: 10.11622/SMEDJ.2020046

Choi, S. W., Gibbons, L. E., and Crane, P. K. (2011). Lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/ itm response theory and monte carlo simulations. *J. Stat. Softw.* 39, 1–30. doi: 10.1002/jcp.22063.Downregulation

Chua, Y. C., Wong, H. H., Abdin, E., Vaingankar, J., Shahwan, S., Cetty, L., et al. (2020). The Recovering Quality of Life 10-item (ReQoL-10) scale in a first-episode psychosis population: validation and implications for patient-reported outcome measures (PROMs). *Early Interv. Psychiatry* 1–9. doi: 10.1111/eip.13050

Crawford, M. J., Robotham, D., Thana, L., Patterson, S., Weaver, T., Barber, R., et al. (2011). Selecting outcome measures in mental health: the views of service users. *J. Ment. Health* 20, 336–346. doi: 10.3109/09638237.2011.577114

DASS (2021). *Depression, Anxiety, Stress Scales (DASS).* Available online at: http://www2.psy.unsw.edu.au/dass/ (accessed April 19, 2021).

Deegan, P. E. (2002). Recovery as a self-directed process of healing and transformation. *Occup. Ther. Ment. Health* 17, 5–21. doi: 10.1300/J004v17n03_02

DeVellis, R. F. (2017). *Scale Development: Theory and Applications*, 4the Edn. Los Angeles: SAGE.

Ellison, M. L., Belanger, L. K., Niles, B. L., Evans, L. C., and Bauer, M. S. (2016). Explication and definition of mental health recovery: a systematic review. *Adm. Policy Ment. Health Ment. Health Serv. Res.* 45, 91–102. doi: 10.1007/s10488-016-0767-9

Ferrando, P. J., and Lorenzo-Seva, U. (2000). Universitat Rovira i Virgili unrestricted versus restricted factor analysis of multidimensional test items: some aspects of the problem and some suggestions. *Psicológica* 21, 301–323.

Fleiss, J. L. (1999). *The Design and Analysis of Clinical Experiments*. New York, NY: Wiley.

Galea, S., Merchant, R. M., and Lurie, N. (2020). The mental health consequences of COVID-19 and physical distancing: the need for prevention and early intervention. *JAMA Intern. Med.* 180:817. doi: 10.1001/jamainternmed.2020.1562

Garner, B. R., Scott, C. K., Dennis, M. L., and Funk, R. R. (2014). The relationship between recovery and health-related quality of life. *J. Subst. Abuse Treat.* 47, 293–298. doi: 10.1016/j.jsat.2014.05.006

Goldhagen, J., Fafard, M., Komatz, K., Eason, T., and Livingood, W. C. (2016). Erratum to: community-based pediatric palliative care for health related quality of life, hospital utilization and costs lessons learned from a pilot study. *BMC Palliat. Care* 15:73. doi: 10.1186/s12904-016-0138-z

Herdman, M., Gudex, C., Lloyd, A., Janssen, M. F., Kind, P., Parkin, D., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual. Life Res.* 20, 1727–1736. doi: 10.1007/s11136-011-9903-x

Hernandez, A., Hidalgo, D., Hambleton, R. K., and Gomez-Benito, J. (2020). International Test Commission guidelines for test adaptation: a criterion checklist/Directrices de la Comisidn International de Test para la adaptation de test: un listado de verification. *Psicothema* 32, 390–398. doi: 10.7334/psicothema2019.306

Keetharuth, A. D., Bjorner, J. B., Barkham, M., Browne, J., Croudace, T., and Brazier, J. (2018a). Exploring the item sets of the Recovering Quality of Life (ReQoL) measures using factor analysis. *Qual. Life Res.* 28, 1005–1015. doi: 10.1007/s11136-018-2091-1

Keetharuth, A. D., Brazier, J. B., Connell, J., Bjorner, J. B., Carlton, T., Buck, L., et al. (2018b). Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. *Br. J. Psychiatry* 212, 42–49. doi: 10.1192/bjp.2017.10

Khan, K. S., Mamun, M. A., Griffiths, M. D., and Ullah, I. (2020). The mental health impact of the COVID-19 pandemic across different cohorts. *Int. J. Ment. Health Addict.* 3, 1–7. doi: 10.1007/s11469-020-00270-8

Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O., and Löwe, B. (2007). Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann. Intern. Med.* 146:317. doi: 10.7326/0003-4819-146-5-200703060-00004

Lamers, L. M., Bouwmans, C. A., van Straten, A., Donker, M. C., and Hakkaart, L. (2006). Comparison of EQ-5D and SF-6D utilities in mental health. *Health Econ.* 15, 1229–1236. doi: 10.1002/hec.1125

Leamy, M., Bird, V., Le Boutillier, C., Williams, J., and Slade, M. (2011). Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *Br. J. Psychiatry* 199, 445–452. doi: 10.1192/bjp.bp.110.083733

Lovibond, S. H., and Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales*, 2nd Edn, Sydney, NSW: Psychology Foundation of Australia.

Mak, W. W. S., Chan, R. C. H., Pang, I. H. Y., Chung, N. Y. L., Yau, S. S. W., and Tang, J. P. S. (2016). Effectiveness of Wellness Recovery Action Planning (WRAP) for Chinese in Hong Kong. *Am. J. Psychiatr. Rehabil.* 19, 235–251. doi: 10.1080/15487768.2016.1197859

Markon, K. E. (2019). Bifactor and hierarchical models: specification, inference, and interpretation. *Annu. Rev. Clin. Psychol.* 15, 51–69.

McDonald, R. P. (1999). *Test Theory?: A Unified Treatment*. Mahwah, NJ: L. Erlbaum Associates.

Mulhern, B., Mukuria, C., Barkham, M., Knapp, M., Byford, S., Soeteman, D., et al. (2014). Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D. *Br. J. Psychiatry* 205, 236–243.

Papaioannou, D., Brazier, J., and Parry, G. (2011). How valid and responsive are generic health status measures, such as EQ-5D and SF-36, in Schizophrenia? A systematic review. *Value Health* 14, 907–920.

Papaioannou, D., Brazier, J., and Parry, G. (2013). How to measure quality of life for cost-effectiveness analyses of personality disorders: a systematic review. *J. Pers. Disord.* 27, 383–401. doi: 10.1521/pedi_2013_27_075

R Core Team (2013). *A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Reise, S. P., Bonifay, W. E., and Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *J. Pers. Assess.* 95, 129–140. doi: 10.1080/00223891.2012.725437

Reise, S. P., Moore, T. M., and Haviland, M. G. (2010). Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *J. Pers. Assess.* 92, 544–559. doi: 10.1080/00223891.2010.496477

ReQoL (2021). *ReQoL: OVERVIEW*. Available online at: https://www.reqol.org.uk/p/overview.html (accessed April 19, 2021).

Roberts, J., Lenton, P., Keetharuth, A. D., and Brazier, J. (2014). Quality of life impact of mental health conditions in England: results from the adult psychiatric morbidity surveys. *Health Qual. Life Outcom.* 12:6. doi: 10.1186/1477-7525-12-6

Robinson, M., Johnson, A. M., Walton, D. M., and MacDermid, J. C. (2019). A comparison of the polytomous Rasch analysis output of RUMM2030 and R (ltm/eRm/TAM/lordif). *BMC Med. Res. Methodol.* 19:36. doi: 10.1186/s12874-019-0680-5

Schönberger, M., and Ponsford, J. (2009). The factor structure of the hospital anxiety and depression scale in individuals with traumatic brain injury. *Psychiatry Res.* 179, 342–349. doi: 10.1016/j.psychres.2009.07.003

Shigemura, J., Ursano, R. J., Morganstein, J. C., Kurosawa, M., and Benedek, D. M. (2020). Public responses to the novel 2019 coronavirus (2019-nCoV) in Japan: mental health consequences and target populations. *Psychiatry Clin. Neurosci.* 74, 281–282. doi: 10.1111/pcn.12988

Spitzer, R. L., Kroenke, K., Williams, J. B., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Intern. Med.* 166, 1092–1097.

Tang, C., Xiong, Y., Wu, H., and Xu, J. (2018). Adaptation and assessments of the Chinese version of the ICECAP-A measurement. *Health Qual. Life Outcom.* 16, 11–45. doi: 10.1186/s12955-018-0865-3

Tong, X., An, D., McGonigal, A., Park, S. P., and Zhou, D. (2015). Validation of the Generalized Anxiety Disorder-7 (GAD-7) among Chinese people with epilepsy. *Epilepsy Res.* 120, 31–36.

van Aken, B. C., de Beurs, E., Mulder, C. L., and van der Feltz-Cornelis, C. M. (2020). The dutch recovering quality of life questionnaire (ReQoL) and its psychometric qualities. *Eur. J. Psychiatry* 34, 99–107.

Wang, K., Shi, H. S., Geng, F. L., Zou, L. Q., Tan, S. P., Wang, Y., et al. (2016). Cross-cultural validation of the depression anxiety stress scale-21 in China. *Psychol. Assess.* 28, e88–e100.

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., et al. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported Outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health* 8, 94–104. doi: 10.1111/j.1524-4733.2005.04054.x

Wong, E. L. Y., Ramos-Goñi, J. M., Cheung, A. W. L., Wong, A. Y. K., and Rivero-Arias, O. (2018). Assessing the use of a feedback module to model EQ-5D-5L health states values in Hong Kong. *Patient* 11, 235–247. doi: 10.1007/s40271-017-0278-0

Wong, E. L. Y., Xu, R. H., and Cheung, A. W. L. (2019). Health-related quality of life among patients with hypertension: population-based survey using EQ-5D-5L in Hong Kong SAR, China. *BMJ Open* 9:e032544. doi: 10.1136/bmjopen-2019-032544

Wong, E. L. Y., Xu, R. H., and Cheung, A. W. L. (2020). Measurement of health-related quality of life in patients with diabetes mellitus using EQ-5D-5L in Hong Kong, China. *Qual. Life Res.* 29, 1913–1921. doi: 10.1186/1477-7525-8-18

Wong, E. L., Cheung, A. W., Wong, A. Y., Xu, R. H., Ramos-Goñi, J. M., and Rivero-Arias, O. (2019). Normative profile of health-related quality of life for Hong Kong general population using preference-based instrument EQ-5D-5L. *Value Health* 22, 916–924. doi: 10.1016/j.jval.2019.02.014

Wouters, E., Booysen, F. R., Ponnet, K., and Baron van Loon, F. (2012). Wording effects and the factor structure of the hospital anxiety & depression scale in HIV/AIDS patients on antiretroviral treatment in south Africa. *PLoS One* 7:e34881. doi: 10.1371/journal.pone.0034881

Xu, R. H., Cheung, A. W. L., and Wong, E. L. (2017). Examining the health-related quality of life using EQ-5D-5L in patients with four kinds of chronic diseases from specialist outpatient clinics in Hong Kong SAR, China. *Patient Prefer Adheren.* 11, 1565–1572. doi: 10.2147/PPA.S143944

Xu, R. H., Cheung, A. W. L., and Wong, E. L. Y. (2018). Development and validation of an instrument to measure patient engagement in Hong Kong Special Administrative Region, China. *Patient Prefer Adheren.* 12, 1667–1675. doi: 10.2147/PPA.S171026

Yin, S., Njai, R. R., Barker, L., Siegel, R. Z., and Liao, Y. (2016). Summarizing health-related quality of life (HRQOL): development and testing of a one-factor model. *Popul. Health Metr.* 14:22. doi: 10.1186/s12963-016-0091-3

Zhang, Y., and Ma, Z. F. (2020). Impact of the COVID-19 pandemic on mental health and quality of life among local residents in liaoning province, China: a cross-sectional study. *Int. J. Environ. Res. Public Health* 17:2381. doi: 10.3390/ijerph17072381

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binaryand Likert-Type (Ordinal) Item Scores.* Ottawa: Directorate of Human Resources Research and Evaluation National Defense.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 26 August 2021
doi: 10.3389/fpsyg.2021.685137

Check for updates

Compulsive Internet Use Scale: Psychometric Properties and Associations With Sleeping Patterns, Mental Health, and Well-Being in Lithuanian Medical Students During the Coronavirus Disease 2019 Pandemic

Egle Milasauskiene[1*], *Julius Burkauskas*[2], *Aurelija Podlipskyte*[2], *Orsolya Király*[3], *Zsolt Demetrovics*[3,4], *Laurynas Ambrasas*[1] *and Vesta Steibliene*[1,2]

[1] *Clinic of Psychiatry, Lithuanian University of Health Sciences, Kaunas, Lithuania,* [2] *Laboratory of Behavioral Medicine, Neuroscience Institute, Lithuanian University of Health Sciences, Palanga, Lithuania,* [3] *Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary,* [4] *Centre of Excellence in Responsible Gaming, University of Gibraltar, Gibraltar, Gibraltar*

OPEN ACCESS

Edited by:
Paola Gremigni,
University of Bologna, Italy

Reviewed by:
Fabio Presaghi,
Sapienza University of Rome, Italy
Tania Moretta,
University of Padova, Italy

****Correspondence:***
Egle Milasauskiene
egle.milasauskiene@lsmuni.lt

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *24 March 2021*
Accepted: *26 July 2021*
Published: *26 August 2021*

Citation:
Milasauskiene E, Burkauskas J, Podlipskyte A, Király O, Demetrovics Z, Ambrasas L and Steibliene V (2021) Compulsive Internet Use Scale: Psychometric Properties and Associations With Sleeping Patterns, Mental Health, and Well-Being in Lithuanian Medical Students During the Coronavirus Disease 2019 Pandemic. Front. Psychol. 12:685137. doi: 10.3389/fpsyg.2021.685137

Background: The increase in problematic Internet use (PIU) among medical students and resident doctors during the coronavirus disease 2019 (COVID-19) pandemic may be leading to significant impairments in everyday functioning, including sleeping patterns, anxiety, depressive symptoms, and overall well-being. The Compulsive Internet Use Scale (CIUS) has been developed to assess the severity of PIU, however, it has not been elucidated whether this scale is also applicable to medical students and resident doctors. The first aim of this study was to explore the psychometric properties of the Lithuanian version of the CIUS. The second aim was to examine associations between subjectively reported mental health symptoms and PIU during the COVID-19 pandemic.

Methods: A total of 524 medical students and resident doctors (78.60% women, mean age 24 [SD 3] years old) participated in an online survey between December 2020 and February 2021. Participants completed the CIUS, the Pittsburgh Sleep Quality Index (PSQI) questionnaire, the Patient Health Questionnaire-9 (PHQ-9), the Generalized Anxiety Disorder Assessment-7 (GAD-7), and the WHO—Five Well-Being Index questionnaire (WHO-5).

Results: The confirmatory factor analysis (CFA) suggested brief versions (CIUS-5, CIUS-7, and CIUS-9) rather than the original (CIUS-14) version of the CIUS questionnaire as reliable and structurally stable instruments that can be used to measure compulsive Internet use severity in the sample of medical students and resident doctors. The most prevalent online behaviors were social media use (90.1%), online shopping (15.6%), and online gaming/gambling (11.3%). Students with higher CIUS scores reported significantly lower academic achievements during the 6 months ($r = 0.12$–0.13; $p < 0.006$), as well as more severe depressive and anxiety symptoms, worsened sleep quality, and lower sense of well-being ($r = 0.21$–0.41; p's < 0.001). Both, during workdays ($d = 0.87$) and weekend ($d = 0.33$), students spent more time online than resident doctors (p's < 0.001).

Conclusion: The brief, 5-, 7-, and 9-item versions of the Lithuanian CIUS are reliable and valid self-report screening instruments for evaluating the severity of PIU symptoms among the medical student population. Symptoms of PIU during the COVID-19 period were associated with worsened self-reported mental health and everyday functioning.

Keywords: problematic internet use, validation, psychometrics, sleep, anxiety, depression, well-being, COVID-19

INTRODUCTION

During the last few decades, Internet use has not only grown but has also transformed from a tool to collect and share information to a way of connecting with others and the world around us. In January 2021, 4.66 billion people were active Internet users, which means that more than 59% of the global population is currently connected to the Internet (Statista, 2021). Individuals aged 18–24 years old account for 18% of global Internet users and those aged from 25 to 34 account for 32% of global Internet users (Statista, 2021). According to a 2019 report from the Lithuanian National Department of Statistics, the daily usage of Internet among those aged 16–24 years old was 98% and those aged 25–34 years old was 95% (Lithuanian National Department of Statistics, 2019).

The general rise in Internet usage also poses a risk of online activities becoming excessive such as gaming/gambling, streaming, watching pornography, and shopping (Fineberg et al., 2018). Problematic Internet use (PIU) is defined as the excessive and compulsive use of online activities and services which have addictive potential associated with marked functional impairment (Király et al., 2015; Ioannidis et al., 2016, 2019).

Following the spatial distancing recommendations of World Health Organization (2020) imposed by many governments around the world in response to the coronavirus disease 2019 (COVID-19) pandemic, many international experts have expressed concern over worsening PIU. To address this, a group of international experts has recommended guidelines for diminishing the risks of increased Internet usage, including encouraging individuals to reach out to mental health professionals if they experienced high levels of distress or significant difficulties controlling their Internet use or specific online activities (Király et al., 2020). Risk factors contributing to PIU include well-known risk factors such as age, sex, and mental health problems, as well as new risk factors brought about by the COVID-19 pandemic such as higher distress or diminished coping. As reported in a recent study by Sun et al. (2020), coping behaviors, such as Internet use, alcohol consumption, and smoking, during the COVID-19 pandemic increased the risk for subsequent substance use disorders and Internet addiction (Sun et al., 2020). In recent years, especially during the COVID-19 pandemic, medical students and resident doctors have experienced an unprecedented expansion of Internet use for online learning, work, and communication (Almomani et al., 2021). The development of electronic patient records, online networking, and training of healthcare professionals has provided advantages for information sharing, learning, and decision-making (Gill et al., 2013). Medical students and doctor residents are among the more vulnerable population because they have to study and work long day and night hours, in most cases without a structured time schedule. At the same time, they have high inspiration for professional development and unlimited access to use Internet-based technologies. However, despite numerous advantages of Internet-based technologies in medical training and the healthcare system, there is still a growing body of evidence that excessive Internet use could lead to PIU among healthcare professionals and cause comorbid problems related to mental health. The prevalence rate for PIU was reported at 7.2% in the general population (Pan et al., 2020) compared with 9.7% among healthcare professionals (Buneviciene and Bunevicius, 2020) and 30.1% among medical students (Zhang et al., 2018). It is well-recognized that medical students and resident doctors are at a greater risk for mental health problems when compared with the general population (Puthran et al., 2016; Erschens et al., 2019; Zeng et al., 2019; Zhou et al., 2020). Moreover, the sudden shift from in-person learning to online education among the medical student population during the COVID-19 pandemic period was associated with higher odds of having generalized anxiety and being depressed (Essangri et al., 2020; Moitra et al., 2021; Nishimura et al., 2021). Studies also suggest that for resident doctors, excessive Internet use might emerge as a compensatory coping behavior for mental health problems (Ueno et al., 2020). A study examining the correlation between PIU, mental health, and sleep quality among Iranian medical students found that PIU was directly and positively connected to depression ($r = 0.44$; $p < 0.001$), anxiety ($r = 0.45$; $p < 0.001$), and stress ($r = 0.40$; $p < 0.001$) (Shadzi et al., 2020). Furthermore, in their meta-analysis of 22,778 resident doctors, Low et al. (2019) found that the aggregate prevalence of burnout was 51.0%. As a result, mental health problems often severely affect academic performance, dropout rates, and professional development (Dyrbye et al., 2006; Khan et al., 2016; Taha et al., 2019). PIU has been linked not only with sleep problems (You et al., 2021) but also with suicidal behavior mediated by sleep disturbances (Guo et al., 2018).

Reliable and validated tools would allow better identification of the problematic behavior among future medical professionals. However, in Lithuania, a scarcity of available instruments remains for investigating PIU. Most of the existing studies conducted in Lithuania focused on children and teenagers (Blinka et al., 2015; Škarupová et al., 2015; Ustinavičiene et al., 2016), and only two investigated PIU symptoms among students using a psychometrically sound nine-item PIU questionnaire, namely, the PIUQ-9 (Burkauskas et al., 2020; Gecaite-Stonciene et al., 2021). Consequentially, it is of crucial importance to update the toolbox of PIU measurements to investigate the symptomatology in specific samples of individuals who might be particularly vulnerable to the phenomena.

The Compulsive Internet Use Scale (CIUS) has been proposed as a tool to assess the severity of Internet addiction and compulsive, pathological, or problematic online behaviors. In the review conducted by Laconi et al. (2014), over 40 scales were identified as measuring PIU at the time. However, the CIUS was regarded as one of the shortest questionnaires in comparison with other scales measuring PIU severity, while having good psychometric properties at the same time. Thus, this scale might be of particular use for screening medical professionals for PIU as it takes about 3–5 min to complete it.

The CIUS was design based on the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) criteria for Dependence and Pathological Gambling and also includes general features of behavioral addictions which were added based on the recommendations by Griffiths (1999) and the results of a qualitative study executed by the author of the scale (Meerkerk et al., 2003). The scale includes typical symptoms of compulsive Internet use such as the inability to control Internet use, mental and behavioral preoccupation with online activities, agitation associated with the inability to go online, mood change, and conflicts with significant others over Internet use (Meerkerk et al., 2009).

The CIUS has been adapted and psychometrically tested in various languages and specific groups of individuals showing good psychometric properties (Dhir et al., 2015; Lopez-Fernandez et al., 2019). However, there is still debate on the CIUS factor structure, with some studies proposing the CIUS as a one-dimensional instrument (Khazaal et al., 2012; Guertler et al., 2014; Lopez-Fernandez et al., 2019) and others finding better compatibility for a three-factor model consisting of CIUS—Absorption, CIUS—Priorities, and CIUS—Mood regulation (Alavi et al., 2011; Yong et al., 2017). Furthermore, a study by Lopez-Fernandez et al. (2019) found that short versions of the instrument (CIUS-5, CIUS-7, and CIUS-9) provide an even better one-factor structure than the original 14-item questionnaire. To date, the CIUS psychometric properties have not been analyzed in Lithuania among the populations of medical students and resident doctors.

Thus, the current study aimed to examine the psychometric properties of the Lithuanian version of CIUS in a sample of medical students and resident doctors. An additional aim was to examine the relationship between PIU and mental health problems, such as subjective sleep quality, depression and anxiety symptoms, and overall well-being.

MATERIALS AND METHODS

Study Procedure

This cross-sectional study was conducted in the Lithuanian University of Health Sciences (LUHS) departments between December 2020 and February 2021 (during the COVID-19 pandemic). Using the official university mailing system and social media groups for medical students and resident doctors, study participants were invited to fill out an online survey available through Google Forms. Before starting the survey, participants had to provide online informed consent to participate in the study by ticking the appropriate answer "agree/disagree." There were no incentives for study participants on completion.

The survey comprised of scales on compulsive Internet use CIUS, sleeping patterns (Pittsburgh Sleep Quality Index, PSQI), depression symptoms (Patient Health Questionnaire-9, PHQ-9), anxiety symptoms (Generalized Anxiety Disorder Assessment-7, GAD-7), and general well-being (WHO—Five Well-Being Index, WHO-5). The sociodemographic questionnaire was developed by the authors and included data on age, sex, living conditions, family situation, physical activity, academic achievements, participation in academic classes, and Internet use habits of participants including a question whether individuals think that their online behavior is problematic ("yes"/ "no"). Overall, the survey included 62 questions, with a completion time ranging from 18 to 25 min with possibly minimal chances for "survey fatigue." The approval from the Bioethics Committee of the LUHS (No. BEC-LSMU [R]-18) was received and the study was executed in accordance with the principles of the Declaration of Helsinki.

Measures

The CIUS is a self-report 14-item scale for the assessment of the severity of Internet addiction and/or compulsive, pathological, or other Internet use which could be considered as problematic. Each question utilizes a 5-point Likert scale, ranging from 0 ("never") to 4 ("very often") and results in a total score ranging from 0 to 56. A higher score indicates a higher severity of PIU. Shortened versions of the CIUS were proposed by Lopez-Fernandez et al. (2019) (with CIUS-9 including items 1, 3, 4, 5, 7, 9, 11, 12, 14; CIUS-7 including items 1, 3, 5, 7, 9, 11, 12; and CIUS-5 including items 1, 3, 5, 11, 12).

The author of the CIUS approved the use of the scale within the current study (Meerkerk et al., 2009). A double back-translation procedure was used to translate the CIUS questions from English to Lithuanian. An experienced psychiatrist with the knowledge of PIU terminology (VS) performed the original translation from English to Lithuanian. Bilingual (English and Lithuanian) native speakers from the registered academic translation service provider conducted the back translation from Lithuanian to English. Significant differences in the back translation as compared with the original CIUS version were discussed to reach the final consensus.

Sleep patterns of participants were rated using the PSQI, which is a self-rated questionnaire for the assessment of subjective sleep quality over 1 month (Buysse et al., 1989). The PSQI consists of 19 questions each ranging from 0 to 3 points, where 0 points indicate "no difficulties" and 3 points indicate "severe sleep problems." Answers are summed into seven components of a global PSQI: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medication, and daytime dysfunction. The global PSQI score is the sum of the seven components and ranges from 0 to 21. The Lithuanian version of the PSQI (Varoneckas, 2003) showed acceptable internal consistency in this study sample (Cronbach's α was 0.71).

A brief self-rated questionnaire, the PHQ-9 was used for measuring the severity of depressive symptoms (Kroenke et al., 2001). Nine items of the questionnaire are based on the depression diagnostic criteria of the DSM-IV; each of the nine items scores from 0 ("Not at all") to 3 ("Nearly every day"). The total score equals the sum of each of the nine items and ranges from 0 to 27, where the higher scores indicate more severe depressive symptoms. Internal consistency of the scale in this study sample is considered good (Cronbach's α 0.84).

The GAD-7 was used to measure the severity of anxiety symptoms (Spitzer et al., 2006). In this seven-item questionnaire, participants have to provide their responses on a 4-point Likert scale ranging from 0 ("Not at all") to 3 ("Nearly every day"). The total score is the sum of all seven answers and ranges from 0 to 21; the higher sum score indicates more severe anxiety symptoms. Internal consistency of the scale in our study sample is considered good (Cronbach's α was 0.91).

Current subjective psychological well-being was assessed using the WHO-5 (World Health Organization, 1998; Topp et al., 2015). This instrument consists of five statements about the period of the past 2 weeks and responses range from 0 ("At no time") to 5 ("All of the time"). The sum score ranges from 0 to 25. Because scales measuring health-related quality of life are conventionally translated to a percentage scale from 0 (absent) to 100 (maximal), it is recommended to multiply the sum score by four to give the final index: 0 means the worst well-being and 100—the best imaginable well-being. Internal consistency of the scale is in this study sample is considered good (Cronbach's α was 0.88).

Statistical Analysis

Descriptive statistics, internal consistency (Cronbach's α), and exploratory and confirmatory factor analyses (EFA and CFA, respectively), were employed to analyze the reliability and the structural validity of the questionnaire. EFAs were performed to analyze the factor structure of the CIUS by using a common factor method including oblique Promax rotation. Principal axis factoring involved data structure analysis that examined shared variances that are considered unique to individual measurement. Items with a factor loading of 0.3 were considered acceptable (Streiner et al., 2015). The test of sphericity suggested by Bartlett was used to determine whether the correlation between the items was adequate for conducting the factor analysis, with a $p < 0.05$ indicating suitability for structure detection (Tabachnick et al., 2007). The Kaiser–Meyer–Olkin (KMO) index was used to test the sampling adequacy (KMO index > 0.6) (Tabachnick et al., 2007). A scree plot was used for the factor pattern interpretation to decide on the number of the CIUS factors.

Later, CFA was performed for the validation of the factor structure considered in EFA using the maximum likelihood method. In this study, we planned to compare the fit indices of the one- and three-factor models and the proposed one-factor solution of the shortened versions of the instrument (i.e., CIUS-9, CIUS-7, and CIUS-5). The Analysis of Moment Structures (AMOS) 22.0 software was used to test the one-factor model of CIUS-14, the CIUS-9, the CIUS-7, and the CIUS-5, and the three-factor model of CIUS-14 by CFA.

Proposed thresholds for the CFA fit indices were: comparative fit index (CFI) > 0.90 adequate and > 0.95 good, Tucker–Lewis Index (TLI) > 0.90 adequate and > 0.95 good, normed fit index (NFI) > 0.90 adequate and > 0.95 good, root-mean-square error of approximation (RMSEA) < 0.08, $\chi 2$/df with desired range of values 2–5 (Hooper et al., 2008; Brown, 2015).

Finally, using the two-tailed Student's *t*-test and the Fisher's χ^2 test, we compared the CIUS scores and the sociodemographic characteristics and subjective psychological assessments between students and resident doctors. Pearson's correlational analysis was used to assess associations between the CIUS scores and subjective psychological measures. Statistical analyses were performed with the Statistical Package for the Science Software v.22 (SPSS, Chicago, IL, United States). The level of significance was set at $p < 0.05$.

RESULTS

In total, 1,064 resident doctors and 1,650 medical students were invited to participate in the study. Of those invited to participate, significantly more responses were received from the medical student group than the resident doctors (20.8 vs. 16.9%, respectively; $p < 0.05$); and the respondents were younger when compared with non-respondents ($p < 0.05$). The response rate was not associated with the sex of participants. Overall, the response rate was 19.3% with a total of 524 individuals who completed the survey.

Of the 524 study participants, 65.6% were medical students and 34.4% were resident doctors and the overall group means age was 23.7 (3.1) years old (**Table 1**). The majority of the sample was women, living with a partner or family member, married or in a partnership, and non-smokers. The mean length of Internet use was 4.89 (SD = 2.61) h/day on workdays and 5.07 (SD = 2.68) h/day on weekends. The most prevalent online behavior of the participants was social media use, followed by academic uses, online shopping, and online gaming/gambling. There was an inverse correlation between PIU as measured by the CIUS and academic achievements among the whole study group ($r = -0.345$; $p < 0.001$).

Significant differences between the students and resident doctors were observed in time spent online during workdays [$M = 5.59$, $SD = 2.61$ vs. $M = 3.56$, $SD = 2.03$; $t_{(522)} = 9.79$, $p < 0.001$, large effect $d = 0.87$] and weekends [$M = 5.37$, SD = 2.69 vs. $M = 4.51$, SD = 2.56; $t_{(522)} = 3.54$, $p < 0.001$, small effect $d = 0.33$]. However, differences between the aforementioned groups were not significant in the case of the CIUS total score and for other mental health indices as measured by the PHQ-9, the GAD-7, the PSQI index, and the WHO-5 (*p*'s > 0.05).

Psychometric Properties of the Lithuanian Version of the CIUS in the Sample of Medical Students and Doctor Residents

The Lithuanian version of the 14-item CIUS had a very high internal consistency of $\alpha = 0.90$ for the sample of Lithuanian medical students and resident doctors. The total correlation for

TABLE 1 | Sociodemographic characteristics and self-reported psychological assessments in study participants.

Characteristics	All, $n = 524$
Age, years; mean (SD)	23.7 (3.1)
Types of studies, n (%)	
Medical student	344 (65.6)
Doctor resident	180 (34.4)
Sex, n (%)	
Women	412 (78.6)
Men	112 (21.4)
Living condition, n (%)	
Alone	174 (33.2)
With partner/family members	350 (66.8)
Marital state, n (%)	
Single	266 (43.1)
Married/partnership	298 (56.9)
Smoking, n (%)	
Yes	112 (21.4)
No	412 (78.6)
Physical activity, hours/day; mean (SD)	0.80 (0.66)
Time using internet	
Workdays, hours/day; mean (SD)	4.89 (2.61)
Weekends, hours/day; mean (SD)	5.07 (2.68)
The online behavior, n (%)	
Academic	406 (77.5)
Online gaming/gambling	59 (11.3)
Online shopping	82 (15.6)
Online social media	472 (90.1)
Online relationship/partnership	4 (0.8)
Online pornography	6 (1.1)
Streaming (movies/shows etc.)	19 (3.6)
Academic achievements, n (%)	
Excellent/higher than average	228 (43.5)
Lower than average/poor	296 (56.5)
Participation in academic classes, n (%)	
In the most of the classes	454 (86.6)
Only mandatory classes/not participate at all	70 (13.4)
Use of medications to treat depression/anxiety/cope with stress	44 (8.4)
Five Well-Being Index WHO-5, total score; mean (SD)	51.9 (18.9)
Global PSQI index, mean (SD)	6.6 (3.0)
PHQ-9, total score; mean (SD)	9.1 (5.7)
GAD-7, total score; mean (SD)	7.4 (5.1)

SD, standard deviation; WHO-5, World health Organization—Five Well-Being Index; PSQI, Pittsburgh Sleep QualityIndex; PHQ-9, Patient Health Questionnaire-9, GAD-7, Generalized Anxiety Disorder Assessment-7.

all items is shown in **Table 2**. Internal consistency for the three-factor model was good with α ranging from 0.82 to 0.85. Internal consistency of the CIUS-9 and the CIUS-7 was also good with α of 0.84 and 0.81, respectively. Internal consistency of the CIUS-5 was considered acceptable with α of 0.73.

The KMO of 0.901 indicated the applicability of data for the factor analysis. The test of sphericity ($p < 0.001$) suggested by Bartlett supported the factorability of the correlation matrix. EFAs were performed including the solutions of one- and three-factor models (**Table 3**). The factor loadings for the one-factor model ranged from 0.460 to 0.757 and factor loadings for the three-factor model ranged from 0.344 to 0.940. A one-factor model was determined to explain 45.1% and the three-factor model was determined to explain 62.3% of the total variance, which is considered sufficient for a coherent construct of the CIUS. While the scree plot (**Figure 1**) indicated a one-factor structure, results of the EFA suggested both models as acceptable.

Thus, CFAs were then used to confirm the construct dimensionality of the CIUS by testing the solutions of both one-factor and three-factor models. The distribution of the CIUS scores was normal (absolute values skewness from 0.043 to 0.613 and kurtosis from 0.418 to 1.257) (Kim, 2013) and the maximum likelihood estimation method was applied for the CFAs. Both the one-factor model (RMSEA = 0.140, RMSEA 90% CI = 0.131–0.148; CFI = 0.774; NFI = 0.758; TLI = 0.733; $\chi^2/df = 11.2$) and the three-factor model (RMSEA = 0.094, RMSEA 90% CI = 0.085–0.104; CFI = 0.911; NFI = 0.895; TLI = 0.889; $\chi^2/df = 5.67$) showed unacceptable fit. However, we also tested the one-factor solution of the shortened versions of the CIUS as proposed by Lopez-Fernandez et al. (2019). All three versions of the shortened CIUS, namely, the CIUS-9, the CIUS-7, and the CIUS-5 showed acceptable model fit for the one-factor solutions according to the CFAs (**Table 4**).

PIU in Relation to Sleeping Quality, Mental Health, and Well-Being in Lithuanian Medical Students and Resident Doctors

Analysis of the associations among self-reported psychological measures and PIU showed that higher CIUS total scores were associated with more severe depressive and anxiety symptoms, worsened sleep quality, and lower overall well-being (**Table 5**). Answers ("yes"/ "no") to the question on whether participants considered their online behavior as problematic also positively correlated with the CIUS total score (**Table 5**).

DISCUSSION

The main finding of the current study was the confirmation of the use and good psychometric characteristics of the shortened CIUS scales in a population of medical students and resident doctors. Symptoms of PIU were associated with worsened anxiety and depression symptoms, sleep patterns and lowered self-reported sense of well-being.

Results of the internal consistency of the Lithuanian CIUS versions are in line with other studies which found the CIUS-9, CIUS-7, and CIUS-5 internal consistency scores ranging between 0.74 and 0.99 (Lopez-Fernandez et al., 2019). Good psychometric properties of the instrument allow it to be used by health professionals to identify and monitor PIU symptom severity in medical students and resident doctors. Having a psychometrically valid instrument is of crucial importance for the accurate assessment of PIU within vulnerable and at-risk populations. The one-factor model solution for CIUS-9, CIUS-7,

TABLE 2 | Means, robust estimator, and intercorrelations of the compulsive internet use scale items.

Item	Mean (SD)	Huebr's M estimator	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. CIUS total	22.1 (10.5)	22.1	1														
2. CIUS1	2.13 (1.09)	2.17	0.753*	1													
3. CIUS2	2.09 (1.09)	2.15	0.762*	0.729*	1												
4. CIUS3	1.03 (1.13)	0.94	0.516*	0.361*	0.383*	1											
5. CIUS4	1.25 (1.10)	1.19	0.646*	0.453*	0.505*	0.325*	1										
6. CIUS5	2.06 (1.14)	2.11	0.635*	0.455*	0.467*	0.197*	0.395*	1									
7. CIUS6	1.09 (1.08)	1.02	0.688*	0.432*	0.442*	0.343*	0.419*	0.372*	1								
8. CIUS7	0.99 (1.09)	0.91	0.693*	0.435*	0.417*	0.371*	0.441*	0.366*	0.656*	1							
9. CIUS8	2.21 (1.26)	2.27	0.718*	0.575*	0.585*	0.397*	0.305*	0.414*	0.450*	0.419*	1						
10. CIUS9	1.66 (1.19)	1.71	0.725*	0.549*	0.544*	0.305*	0.349*	0.447*	0.461*	0.396*	0.722*	1					
11. CIUS10	0.89 (1.07)	0.99	0.693*	0.395*	0.438*	0.351*	0.438*	0.356*	0.493*	0.532*	0.385*	0.438*	1				
12. CIUS11	0.88 (1.10)	0.90	0.628*	0.412*	0.430*	0.281*	0.411*	0.342*	0.355	0.409*	0.296*	0.338*	0.557*	1			
13. CIUS12	2.49 (1.09)	2.57	0.675*	0.493*	0.479*	0.158*	0.374*	0.430*	0.311*	0.328*	0.452*	0.440*	0.352*	0.362*	1		
14. CIUS13	2.34 (1.26)	2.44	0.601*	0.413*	0.363*	0.121*	0.286*	0.339*	0.259*	0.297*	0.339*	0.376*	0.323*	0.323*	0.740*	1	
15. CIUS14	1.01 (1.10)	0.93	0.626*	0.349*	0.357*	0.249*	0.390*	0.352*	0.487*	0.459*	0.302*	0.360*	0.403*	0.403*	0.360*	0.364*	1

**Correlation is significant at the 0.010 level (two-tailed). CIUS, Compulsive Internet Use Scale.*

TABLE 3 | Factors loadings of each item of the CIUS.

Compulsive internet use scale	One-factor model	Three-factor model		
	F1	F1	F2	F3
1. Do you find it difficult to stop using the Internet when you are online?	0.745		0.638	
2. Do you continue to use the Internet despite your intention to stop?	0.757		0.645	
3. Do others (e.g., partner, children, parents) say you should use the Internet less?	0.406		0.344	
4. Do you prefer to use the Internet instead of spending time with others (e.g., partner, children, parents)?	0.611	0.513		
5. Are you short of sleep because of the Internet?	0.596	-	-	-
6. Do you think about the Internet, even when not online?	0.663	0.682		
7. Do you look forward to your next Internet session?	0.666	0.770		
8. Do you think you should use the Internet less often?	0.692		0.940	
9. Have you unsuccessfully tried to spend less time on the Internet?	0.702		0.732	
10. Do you rush through your (home) work in order to go on the Internet?	0.664	0.737		
11. Do you neglect your daily obligations (work, school, or family life) because you prefer to go on the Internet?	0.588	0.577		
12. Do you go on the Internet when you are feeling down?	0.635			0.860
13. Do you use the Internet to escape from your sorrows or get relief from negative feelings?	0.545			0.829
14. Do you feel restless, frustrated, or irritated when you cannot use the Internet?	0.583	0.639		
Eigen value	6.32	6.32	1.30	1.11
% of variance	45.1	45.1	9.25	7.92
Cumulative % of variance	45.1	45.1	54.4	62.3

and CIUS-5 through CFA complements the analysis performed by Lopez-Fernandez et al. (2019) who also found CIUS-14 in CFA to provide a poor global overall model fit. However, there was confirmatory evidence of the convergent construct validity of CIUS, based on the associations of the total score of questionnaires with the mental health, sleep patterns, and well-being indices.

The majority of the study participants reported an academic purpose for Internet use, however, PIU was inversely associated with academic achievements in this sample. Internet use for entertainment purposes was found to be a significant predictor of Internet addiction among women non-medical students (Abdel-Salam et al., 2019). Ueno et al. (2020) also suggested that immersion into virtual reality could be used as a coping

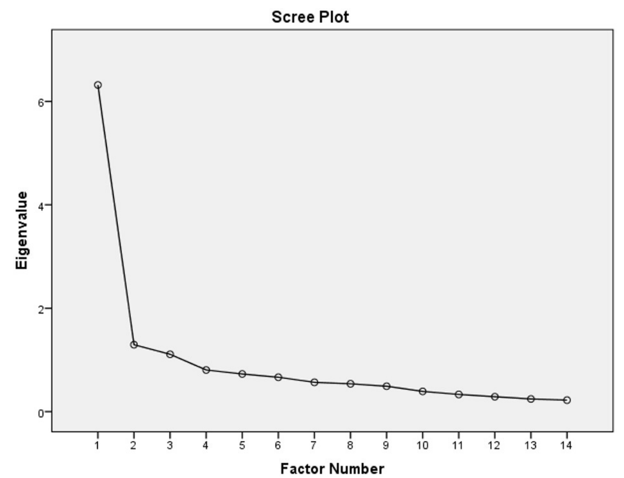

FIGURE 1 | Scree plot illustrating the factor loadings with parallel analysis of the Lithuanian version of the Compulsive Internet Use Scale (CIUS).

mechanism for large academic load and negative psychological states (Ueno et al., 2020).

In contrast, in medical students and resident doctors, PIU severity was associated with several mental health problems, including worse sleep quality, lower subjective feeling of psychological well-being, higher rates of depression, and anxiety compared with medical students and resident doctors without PIU. The effect size of these results varied from small in the case of anxiety, sleep quality, and well-being indices to medium in the case of depression symptoms. A recent study conducted in Lithuania in the general student population showed significant associations between PIU (as measured by the PIUQ-9 questionnaire), anxiety, and depressive symptoms (Gecaite-Stonciene et al., 2021). Similarly, symptoms of anxiety, depression, and sleep disturbances were associated with PIU among the medical population in studies conducted in other countries (Capetillo-Ventura and Juárez-Treviño, 2015; Younes et al., 2016; Grover et al., 2019; Shadzi et al., 2020). Recent studies found that PIU correlates with sleep problems, depression, and anxiety in non-medical populations as well (de Vries and Nakamae, 2018; Khazaal et al., 2021).

The study results suggest a relationship between a higher CIUS total score and lower subjective sleep quality. Kalmbach et al. (2017) suggest that sleep deprivation among physicians-in-training leads to higher medical error rates and may also have a negative effect on the mental health of doctors (Gecaite-Stonciene et al., 2021).

We did not find any significant associations between sex, age, living conditions, marital status, time spent using the Internet, smoking, physical activity, and use of psychotropic medications and PIU, whereas other studies found significant positive correlations between Internet addiction and urban living (Cao et al., 2011; Stavropoulos et al., 2013; Yasuma et al., 2019), and gaming on the Internet (Tsumura and Kanda, 2018) and time spent online (Laconi et al., 2019) among non-medical populations.

While time spent online is an important contributor to PIU, several studies (Mazhari, 2012; Orben and Przybylski, 2019; Coyne et al., 2020), including the one by the CIUS author (Meerkerk et al., 2009), suggest that specific activities online (e.g., gaming or watching pornography) might contribute more to the PIU symptomatology than the actual time spent online. The most prevalent activity online in this study sample was academic search and social media use. Other studies also suggest that time spent on these activities is not directly related to PIU (Mazhari, 2012; Orben and Przybylski, 2019; Coyne et al., 2020). However, this study reveals the negative impact of PIU on both mental health and sleep quality, which is in line with the results of Shadzi et al. (2020).

However, significant differences between the student and resident doctor groups emerged in time spent online. During workdays, students tend to spend ∼2 h more time online than resident doctors. The difference persists during weekends, although the time difference is much lower (∼45 min). This difference could be explained by the roles and responsibilities of the groups, as resident doctors are usually full or part-time workers and have to take care of other duties which are not online both during workdays and weekends.

Even though we were the first to validate a Lithuanian language version of the short CIUS scales and use it in assessing associations with mental health problems among the students of medical faculty and resident doctors in Lithuania, the study should be interpreted in the context of its design and limitations.

The results depended on self-reported data and cannot be generalized as only a small sample of medical students and resident doctors from Lithuania were included. Also, this was a cross-sectional study and causality could not be established. Data collection may have been affected by selection bias, as questionnaires were distributed only through online platforms, so participants who use the Internet more often were possibly more likely to fill out the online questionnaire and were also more likely to be affected by PIU. In this survey, we had not been collecting the COVID-19-related medical status of participants. Living with or treating a person who was (or is) tested positive with COVID-19 or experiencing a loss because of the COVID-19 may have been important stressors (to control) affecting the mental health of participants during COVID-19 and probably the Internet use practices of participants.

We have chosen the CIUS for its brevity and the possibility to capture key symptoms of PIU such as loss of control, preoccupation with online activities, and mood change. As measured with the CIUS, PIU was found to be associated with higher levels of anxiety, depression, sleep problems, lower subjective feeling of psychological well-being, and poorer academic performance. The strongest effect was observed for depression symptoms.

It remains unknown whether PIU predicts subsequent psychiatric disorders or if the causality is reversed. The findings of this study should also be taken into account by the

TABLE 4 | Confirmatory Factor Analyses (CFA) were conducted using the original CIUS.

	α	χ^2/df	df	χ^2	*p*	CFI	TLI	NFI	RMSEA	RMSA 90%CI	p_{close}
F1—CIUS-14	0.904	11.21	77	862,761	<0.001	0.774	0.733	0.758	0.140	0.131–0.148	<0.001
F3—CIUS-14	0.837–0.845	5.67	62	351.237	<0.001	0.911	0.889	0.895	0.094	0.085–0.104	<0.001
F1—CIUS-9	0.844	3.65	27	98.513	<0.001	0.948	0.930	0.930	0.071	0.056–0.087	0.010
F1—CIUS-7	0.807	3.73	14	52.153	<0.001	0.960	0.940	0.947	0.072	0.052–0.094	0.037
F1—CIUS-5	0.727	4.58	5	22.882	<0.001	0.965	0.930	0.956	0.083	0.050–0.118	0.049

F1, one-factor model. F3, a three-factor model (Item 5 is not included due to insufficient factor loading on EFA). CIUS, the Compulsive Internet Use Scale (number represents the sum of items). χ^2, Chi-square; df, degrees of freedom; CFI, comparative fit index; TLI, Tucker–Levis Index; RMSEA, root-mean-square error of approximation; 90% CI, 90% confidence interval of the RMSEA; p_{close}, provides a one-sided test of the null hypothesis that the RMSEA is equal to 0.05 in the population.

TABLE 5 | Pearson's correlations between the CIUS scores and depression, anxiety, subjective sleep, and well-being indices.

	CIUS-9 r (*p*)	CIUS-7 r (*p*)	CIUS-5 r (*p*)
Depression symptoms as measured with the Patient Health Questionnaire-9	0.410 (<0.001)	0.411 (<0.001)	0.410 (<0.001)
Anxiety symptoms as measured with the Generalized Anxiety Disorder Scale 7	0.223 (<0.001)	0.216 (<0.001)	0.213 (<0.001)
Question on whether the participants considered their online behavior as problematic ("yes"/ "no")	0.470 (<0.001)	0.467 (<0.001)	0.443 (<0.001)
Subjective sleep quality			
Subjective sleep quality	0.242 (<0.001)	0.251 (<0.001)	0.254 (<0.001)
Sleep latency	0.118 (0.007)	0.127 (0.003)	0.125 (0.004)
Sleep duration	0.100 (0.022)	0.095 (0.029)	0.113 (0.010)
Habitual sleep efficiency	0.025 (0.566)	0.052 (0.238)	0.057 (0.195)
Sleep disturbances	0.174 (<0.001)	0.163 (<0.001)	0.159 (<0.001)
Use of sleeping medication	0.146 (0.001)	0.140 (0.001)	0.151 (0.001)
Daytime dysfunction	0.334 (<0.001)	0.333 (<0.001)	0.337 (<0.001)
Global Pittsburg Sleep Quality Index, total score	0.271 (<0.001)	0.279 (<0.001)	0.287 (<0.001)
Subjective psychological well-being			
I have felt cheerful in good spirits	−0.237 (<0.001)	−0.233 (<0.001)	−0.235 (<0.001)
I have felt calm and relaxed	−0.174 (<0.001)	−0.188 (<0.001)	−0.181 (<0.001)
I have felt active and vigorous	−0.267 (<0.001)	−0.265 (<0.001)	−0.261 (<0.001)
I woke up feeling fresh and rested	−0.207 (<0.001)	−0.219 (<0.001)	−0.233 (<0.001)
My daily life has been filled with things that interest me	−0.234 (<0.001)	−0.243 (<0.001)	−0.246 (<0.001)
Well-Being Index (WHO-5)	−0.271 (<0.001)	−0.279 (<0.001)	−0.281 (<0.001)

university hospital administrative staff. Despite the numerous advantages of Internet use, if used in a problematic way, Internet use could be associated with poor labor productivity, a higher prevalence of medical errors, and a risk to the safety of patients.

In conclusion, the Lithuanian versions of the CIUS-9, the CIUS-7, and the CIUS-5 are reliable and valid instruments for evaluating the severity of PIU symptoms among the medical student population. In medical students and resident doctors, studying during the COVID-19 pandemic, symptoms of PIU were associated with worsened subjectively evaluated mental health and everyday functioning.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Bioethics Committee of Lithuanian university of Health sciences. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JB and VS designed the study. EM and LA collected and analyzed the data. AP: statistical analyses. EM, JB, LA, AP, ZD, OK, and VS: drafted and edited the manuscript. All authors contributed to the manuscript and approved the final version.

FUNDING

JB has received funding from the Research Council of Lithuania (LMTLT), project agreement No. S-GEV-20-5.

The contribution of ZD was supported by the Hungarian National Research, Development, and Innovation Office (KKP126835; K134807, K128614, ELTE Thematic Excellence Programme 2020, KP2020-IKA-05). OK was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and by the ÚNKP New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development, and Innovation Fund.

ACKNOWLEDGMENTS

We would like to express our deepest gratitude to the students who completed our survey. Our thanks to the author of CIUS Professor Gert-Jan Meerkerk for the permission to use this scale in our study. We thank the COST Action CA16207 European Network for Problematic Usage of the Internet, supported by COST (European Cooperation in Science and Technology: www.cost.eu) for the inspiration to explore this topic.

REFERENCES

Abdel-Salam, D. M., Alrowaili, H. I., Albedaiwi, H. K., Alessa, A. I., and Alfayyadh, H. A. (2019). Prevalence of internet addiction and its associated factors among female students at Jouf University, Saudi Arabia. *J. Egypt. Public Health Assoc.* 94:12. doi: 10.1186/s42506-019-0009-6

Alavi, S. S., Jannatifard, F., Eslami, M., and Rezapour, H. (2011). Validity, reliability and factor analysis of compulsive internet use scale in students of Isfahan's universities. *Health Inform. Manage.* 7, 715–724. Available online at: https://www.sid.ir/en/journal/ViewPaper.aspx?id=201135

Almomani, E. Y., Qablan, A. M., Atrooz, F. Y., Almomany, A. M., Hajjo, R. M., and Almomani, H. Y. (2021). The influence of coronavirus diseases 2019 (COVID-19) pandemic and the quarantine practices on university students' beliefs about the online learning experience in Jordan. *Front. Public Health*. (2021) 8:595874. doi: 10.3389/fpubh.2020.595874

Blinka, L., Škarupová, K., Ševčíková A., Wölfling, K., Müller, K. W., and Dreier, M. (2015). Excessive internet use in European adolescents: What determines differences in severity? *Int. J. Public Health*. 60, 249–256. doi: 10.1007/s00038-014-0635-x

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. New York, NY; London: Guilford Publications.

Buneviciene, I., and Bunevicius, A. (2020). Prevalence of internet addiction in healthcare professionals: Systematic review and meta-analysis. *Int. J. Soc. Psychiatry* 67, 483–491. doi: 10.1177/0020764020959093

Burkauskas, J., Király, O., Demetrovics, Z., Podlipskyte, A., and Steibliene, V. (2020). Psychometric properties of the nine-item problematic internet use questionnaire (PIUQ-9) in a lithuanian sample of students. *Front. Psychiatry* 11:1279. doi: 10.3389/fpsyt.2020.565769

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., and Kupfer, D. J. (1989). The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 28, 193–213. doi: 10.1016/0165-1781(89)90047-4

Cao, H., Sun, Y., Wan, Y., Hao, J., and Tao, F. (2011). Problematic internet use in Chinese adolescents and its relation to psychosomatic symptoms and life satisfaction. *BMC Public Health* 11:802. doi: 10.1186/1471-2458-11-802

Capetillo-Ventura, N., and Juárez-Treviño, M. (2015). Internet addiction in university medical students. *Med. Universitaria* 17, 88–93. doi: 10.1016/j.rmu.2015.02.003

Coyne, S. M., Rogers, A. A., Zurcher, J. D., Stockdale, L., and Booth, M. (2020). Does time spent using social media impact mental health?: An eight year longitudinal study. *Comp. Hum. Behav.* 104:106160. doi: 10.1016/j.chb.2019.106160

de Vries, H. T., and Nakamae, T. (2018). Problematic internet use and psychiatric co-morbidity in a population of Japanese adult psychiatric patients. *BMC Psychiatry* 18:9. doi: 10.1186/s12888-018-1588-z

Dhir, A., Chen, S., and Nieminen, M. (2015). A repeat cross-sectional analysis of the psychometric properties of the Compulsive Internet Use Scale (CIUS) with adolescents from public and private schools. *Comp. Educ.* 86, 172–181. doi: 10.1016/j.compedu.2015.03.011

Dyrbye, L. N., Thomas, M. R., and Shanafelt, T. D. (2006). Systematic review of depression, anxiety, and other indicators of psychological distress among U.S. and Canadian medical students. *Acad Med.* 81, 354–373. doi: 10.1097/00001888-200604000-00009

Erschens, R., Keifenheim, K. E., Herrmann-Werner, A., Loda, T., Schwille-Kiuntke, J., Bugaj, T. J., et al. (2019). Professional burnout among medical students: Systematic literature review and meta-analysis. *Med. Teacher* 41, 172–183. doi: 10.1080/0142159X.2018.1457213

Essangri, H., Sabir, M., Benkabbou, A., Majbar, M. A., Amrani, L., Ghannam, A., et al. (2020). Predictive factors for impaired mental health among medical students during the early stage of the COVID-19 pandemic in Morocco. *Am. J. Trop. Med. Hygiene* 104, 95–102. doi: 10.4269/ajtmh.20-1302

Fineberg, N. A., Demetrovics, Z., Stein, D. J., Ioannidis, K., Potenza, M. N., Grunblatt, E., et al. (2018). Manifesto for a European research network into Problematic Usage of the Internet. *Eur. Neuropsychopharmacol.* 28, 1232–1246. doi: 10.1016/j.euroneuro.2018.08.004

Gecaite-Stonciene, J., Saudargiene, A., Pranckeviciene, A., Liaugaudaite, V., Griskova-Bulanova, I., Simkute, D., et al. (2021). Impulsivity mediates associations between problematic internet use, anxiety, and depressive symptoms in students: a cross-sectional COVID-19 study. *Front. Psychiatry* 12:634464. doi: 10.3389/fpsyt.2021.634464

Gill, H. K., Gill, N., and Young, S. D. (2013). Online technologies for health information and education: a literature review. *J Consumer Health Internet* 17, 139–150. doi: 10.1080/15398285.2013.780542

Griffiths, M. (1999). Internet addiction: fact or fiction? *Psychologist* 12, 246–250.

Grover, S., Sahoo, S., Bhalla, A., and Avasthi, A. (2019). Problematic internet use and its correlates among resident doctors of a tertiary care hospital of North India: A cross-sectional study. *Asian J Psychiatry* 39, 42–47. doi: 10.1016/j.ajp.2018.11.018

Guertler, D., Broda, A., Bischof, A., Kastirke, N., Meerkerk, G. J., John, U., et al. (2014). Factor structure of the compulsive internet use scale. *Cyberpsychol. Behav. Soc. Network*. 17, 46–51. doi: 10.1089/cyber.2013.0076

Guo, L., Luo, M., Wang, W., Huang, G., Xu, Y., Gao, X., et al. (2018). Association between problematic Internet use, sleep disturbance, and suicidal behavior in Chinese adolescents. *J. Behav. Addict.* 7, 965–975. doi: 10.1556/2006.7.2018.115

Hooper, D., Coughlan, J., and Mullen, M. R. (2008). Structural equation modelling: guidelines for determining model fit. *Electr. J. Business Res. Methods* 6, 53–59. doi: 10.21427/D7CF7R

Ioannidis, K., Chamberlain, S. R., Treder, M. S., Kiraly, F., Leppink, E. W., Redden, S. A., et al. (2016). Problematic internet use (PIU): associations with the impulsive-compulsive spectrum. An application of machine learning in psychiatry. *J. Psychiatr. Res.* 83, 94–102. doi: 10.1016/j.jpsychires.2016.08.010

Ioannidis, K., Hook, R., Goudriaan, A. E., Vlies, S., Fineberg, N. A., Grant, J. E., et al. (2019). Cognitive deficits in problematic internet use: meta-analysis of 40 studies. *Br. J. Psychiatry* 215, 639–646. doi: 10.1192/bjp.2019.3

Kalmbach, D. A., Arnedt, J. T., Song, P. X., Guille, C., and Sen, S. (2017). Sleep disturbance and short sleep as risk factors for depression and perceived medical errors in first-year residents. *Sleep* 40:zsw073. doi: 10.1093/sleep/zsw073

Khan, M. A., Alvi, A. A., Shabbir, F., and Rajput, T. A. (2016). Effect of internet addiction on academic performance of medical students. *J. Islam. Int. College* 11, 48–51. Available online at: http://jiimc.riphah.edu.pk/wp-content/uploads/2015/12/Muhammad-Alamgir-Khan.pdf

Khazaal, Y., Chatton, A., Horn, A., Achab, S., Thorens, G., Zullino, D., et al. (2012). French validation of the compulsive internet use scale (CIUS). *Psychiatr Q*. 83, 397–405. doi: 10.1007/s11126-012-9210-x

Khazaal, Y., Chatton, A., Rochat, L., Hede, V., Viswasam, K., Penzenstadler, L., et al. (2021). Compulsive health-related internet use and cyberchondria. *Eur. Addict. Res*. 27, 58–66. doi: 10.1159/000510922

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry Endodont*. 38, 52–54. doi: 10.5395/rde.2013.38.1.52

Király, O., Griffiths, M. D., and Demetrovics, Z. (2015). Internet gaming disorder and the DSM-5: conceptualization, debates, and controversies. *Curr. Addict. Rep*. 2, 254–262. doi: 10.1007/s40429-015-0066-7

Király, O., Potenza, M. N., Stein, D. J., King, D. L., Hodgins, D. C., Saunders, J. B., et al. (2020). Preventing problematic internet use during the COVID-19 pandemic: Consensus guidance. *Comprehen. Psychiatry* 100:152180. doi: 10.1016/j.comppsych.2020.152180

Kroenke, K., Spitzer, R. L., and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med*. 16, 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x

Laconi, S., Rodgers, R. F., and Chabrol, H. (2014). The measurement of Internet addiction: A critical review of existing scales and their psychometric properties. *Comput. Hum. Behav*. 41, 190–202. doi: 10.1016/j.chb.2014.09.026

Laconi, S., Urbán, R., Kaliszewska-Czeremska, K., Kuss, D. J., Gnisci, A., Sergi, I., et al. (2019). Psychometric evaluation of the nine-item problematic internet use questionnaire (PIUQ-9) in nine european samples of internet users. *Front. Psychiatry* (2019) 10:136. doi: 10.3389/fpsyt.2019.00136

Lithuanian National Department of Statistics (2019). *Statistical Yearbook of Lithuania (edition 2019). Information and Communication Technologies*. Available online at: https://osp.stat.gov.lt/lietuvos-statistikos-metrastis/lsm-2019/mokslas-ir-technologijos/informacines-technologijos (accessed March 21, 2021).

Lopez-Fernandez, O., Griffiths, M. D., Kuss, D. J., Dawes, C., Pontes, H. M., Justice, L., et al. (2019). Cross-cultural validation of the compulsive internet use scale in four forms and eight languages. *Cyberpsychol. Behav. Soc. Networ*. 22, 451–464. doi: 10.1089/cyber.2018.0731

Low, Z. X., Yeo, K. A., Sharma, V. K., Leung, G. K., McIntyre, R. S., Guerrero, A., et al. (2019). Prevalence of burnout in medical and surgical residents: a meta-analysis. *Int. J. Environ. Res. Public Health* 16:1479. doi: 10.3390/ijerph16091479

Mazhari, S. (2012). The prevalence of problematic internet use and the related factors in medical students, Kerman, Iran. *Addict. Health* 4:87.

Meerkerk, G., Laluan, A., and van den Eijnden, R. (2003). *Internetverslaving: Hoax of Serieuze Bedreiging Voorde Geestelijke volksgezondheid?[Internet Addiction: Hoax or a Serious Threath to Mental Health]*. Rotterdam: Instituut voor Onderzoek naar Leefwijzen &Verslaving (IVO).

Meerkerk, G. J., Van Den Eijnden, R. J., Vermulst, A. A., and Garretsen, H. F. (2009). The compulsive internet use scale (CIUS): some psychometric properties. *Cyberpsychol. Behav*. 12, 1–6. doi: 10.1089/cpb.2008.0181

Moitra, M., Rahman, M., Collins, P. Y., Gohar, F., Weaver, M., Kinuthia, J., et al. (2021). Mental health consequences for healthcare workers during the COVID-19 pandemic: a scoping review to draw lessons for LMICs. *Front. Psychiatry* 12:22. doi: 10.3389/fpsyt.2021.602614

Nishimura, Y., Ochi, K., Tokumasu, K., Obika, M., Hagiya, H., Kataoka, H., et al. (2021). Impact of the COVID-19 pandemic on the psychological distress of medical students in Japan: cross-sectional survey study. *J. Med. Internet Res.* 23:e25232. doi: 10.2196/25232

Orben, A., and Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nat. Hum. Behav.* 3, 173–182. doi: 10.1038/s41562-018-0506-1

Pan, Y. C., Chiu, Y. C., and Lin, Y. H. (2020). Systematic review and meta-analysis of epidemiology of internet addiction. *Neurosci. Biobehav. Rev.* 118, 612–622. doi: 10.1016/j.neubiorev.2020.08.013

Puthran, R., Zhang, M. W., Tam, W. W., and Ho, R. C. (2016). Prevalence of depression amongst medical students: a meta-analysis. *Med. Educ*. 50, 456–468. doi: 10.1111/medu.12962

Shadzi, M. R., Salehi, A., and Vardanjani, H. M. (2020). Problematic internet use, mental health, and sleep quality among medical students: a path-analytic model. *Indian J. Psychol. Med*. 42, 128–135. doi: 10.4103/IJPSYM.IJPSYM_238_19

Škarupová, K., Ólafsson, K., and Blinka, L. (2015). Excessive internet use and its association with negative experiences: quasi-validation of a short scale in 25 European countries. *Comp. Hum. Behav*. 53, 118–123. doi: 10.1016/j.chb.2015.06.047

Spitzer, R. L., Kroenke, K., Williams, J. B., and Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Inter. Med*. 166, 1092–1097. doi: 10.1001/archinte.166.10.1092

Statista (2021). Available online at: https://www.statista.com/statistics/617136/digital-population-worldwide/ (accessed March 21, 2021).

Stavropoulos, V., Alexandraki, K., and Motti-Stefanidi, F. (2013). Recognizing internet addiction: prevalence and relationship to academic achievement in adolescents enrolled in urban and rural Greek high schools. *J. Adolesc*. 36, 565–576. doi: 10.1016/j.adolescence.2013.03.008

Streiner, D. L., Norman, G. R., and Cairney, J. (2015). *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford: Oxford University Press. doi: 10.1093/med/9780199685219.003.0001

Sun, Y., Li, Y., Bao, Y., Meng, S., Sun, Y., Schumann, G., et al. (2020). Brief report: increased addictive internet and substance use behavior during the COVID-19 pandemic in China. *Am. J. Addict*. 29, 268–270. doi: 10.1111/ajad.13066

Tabachnick, B. G., Fidell, L. S., and Ullman, J. B. (2007). *Using Multivariate Statistics*. Boston, MA: Pearson.

Taha, M. H., Shehzad, K., Alamro, A. S., and Wadi, M. (2019). Internet use and addiction among medical students in Qassim University, Saudi Arabia. *Sultan Qaboos University Med. J*. 19, e142–e147. doi: 10.18295/squmj.2019.19.02.010

Topp, C. W., Østergaard, S. D., Søndergaard, S., and Bech, P. (2015). The WHO-5 well-being index: a systematic review of the literature. *Psychother. Psychosomat*. 84, 167–176. doi: 10.1159/000376585

Tsumura, H., and Kanda, H. (2018). Prevalence and risk factors of internet addiction among employed adults in Japan. *J. Epidemiol*. 28, 202–206. doi: 10.2188/jea.JE20160185

Ueno, T., Ito, K., Murai, T., and Fujiwara, H. (2020). Mental health problems and their association with internet use in medical residents. *Front. Public Health*. 8:587390. doi: 10.3389/fpubh.2020.587390

Ustinavičiene R., Škemiene L., Lukšiene D, Radišauskas, R., Kaliniene G., and Vasilavičius, P. (2016). Problematic computer game use as expression of Internet addiction and its association with self-rated health in the Lithuanian adolescent population. *Medicina* 52, 199–204. doi: 10.1016/j.medici.2016.04.002

Varoneckas, G. (2003). Subjektyvus miego ivertinimas pagal Pitsburgo miego kokybes indeksa. (Subjective evaluation of sleep according to the Pittsburgh Sleep Quality Index). *Nervu ir psichikos ligos*. 4, 31–33.

World Health Organization (1998). *Wellbeing Measures in Primary Health Care/The Depcare Project*. Copenhagen: WHO Regional Office for Europe.

World Health Organization (2020). *Rational Use of Personal Protective Equipment for COVID-19 and Considerations During Severe Shortages: Interim Guidance, 23 December 2020. World Health Organization*. Available online at: https://apps.who.int/iris/handle/10665/338033

Yasuma, N., Watanabe, K., Nishi, D., Ishikawa, H., Tachimori, H., Takeshima, T., et al. (2019). Urbanization and Internet addiction in a nationally representative sample of adult community residents in Japan: A cross-sectional, multilevel study. *Psychiatry Res*. 273, 699–705. doi: 10.1016/j.psychres.2019.01.094

Yong, R. K. F., Inoue, A., and Kawakami, N. (2017). The validity and psychometric properties of the Japanese version of the Compulsive Internet Use Scale (CIUS). *BMC Psychiatry* 17:201. doi: 10.1186/s12888-017-1364-5

You, Z., Mei, W., Ye, N., Zhang, L., and Andrasik, F. (2021). Mediating effects of rumination and bedtime procrastination on the relationship between Internet addiction and poor sleep quality. *J. Behav. Addict.* 9, 1002–1010. doi: 10.1556/2006.2020.00104

Younes, F., Halawi, G., Jabbour, H., El Osta, N., Karam, L., Hajj, A., et al. (2016). Internet addiction and relationships with insomnia, anxiety, depression, stress and self-esteem in university students: a cross-sectional designed study. *PLoS ONE* 11:e0161126. doi: 10.1371/journal.pone.0161126

Zeng, W., Chen, R., Wang, X., Zhang, Q., and Deng, W. (2019). Prevalence of mental health problems among medical students in China: A meta-analysis. *Medicine* 98:e15337. doi: 10.1097/MD.0000000000015337

Zhang, M. W. B., Lim, R. B. C., Lee, C., and Ho, R. C. M. (2018). Prevalence of internet addiction in medical students: a meta-analysis. *Acad. Psychiatry*. 42, 88–93. doi: 10.1007/s40596-017-0794-1

Zhou, A. Y., Panagioti, M., Esmail, A., Agius, R., Van Tongeren, M., and Bower, P. (2020). Factors associated with burnout and stress in trainee physicians: a systematic review and meta-analysis. *JAMA Netw Open*. 3:e2013761. doi: 10.1001/jamanetworkopen.2020.13761

Conflict of Interest: For the past several years, JB has been serving as a consultant to Cogstate, Ltd. VS reported being a consultant to SignantHealth and received personal fees from Lundbeck, Sanofi, Servier, Johnson & Johnson, KRKA, and grants from the Research Council of Lithuania.

The remaining authors declare no commercial or financial relationships could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 September 2021
doi: 10.3389/fpsyg.2021.725344

Check for updates

Development and Validation of the Robust - Pandemic Coping Scale (R-PCS)

*Roberto Burro[1], Giada Vicentini[1], Emmanuela Rocca[1], Veronica Barnaba[1], Rob Hall[2,3] and Daniela Raccanello[1]**

[1]Department of Human Sciences, University of Verona, Verona, Italy, [2]Environmetrics Pty Ltd., Killara, NSW, Australia, [3]Department of Psychology, Macquarie University, Sydney, NSW, Australia

The psychological consequences of epidemics/pandemics, such as the COVID-19 pandemic, include an increase in psychopathological symptoms, such as depression, anxiety, and stress, and negative emotions, such as fear. However, relatively little attention has been paid to how people cope with the pandemic. Coping is a multi-component process, helping to diminish the traumatic impact of stressful events in a variety of ways. We studied how university students coped with the first wave of the COVID-19 pandemic, by developing the Robust - Pandemic Coping Scale (R-PCS), a new scale for measuring coping strategies related to epidemics/pandemics. The scale is based on a classification of coping strategies referred to the needs of competence, relatedness, and autonomy. To create a robust scale, such that the item values would be independent of the sample used for developing it, we employed Rasch modeling. We used a sample of 2,987 Italian university students who participated in an online survey including the R-PCS and the Power to Live with Disasters Questionnaire (PLDQ), during March 2020. First, we applied a dual approach combining exploratory and confirmatory factor analyses, which supported the goodness of a 4-factor model (i.e., Despair, Adjustment, Proactivity, and Aversion) for the R-PCS, invariant across gender and age of respondents (younger or as old as 23 years, older than 23 years). We then transformed the raw scores of the R-PCS into interval logit scale scores applying the Rasch model. Second, our findings supported the discriminant validity and the criterion validity of the R-PCS, examining the correlations with the PLDQ. They also confirmed its predictive validity: the R-PCS scores were related to 2-month-later enjoyment and anger, indicating that Adjustment and Proactivity were adaptive while Despair and Aversion were maladaptive. Third, our study revealed gender and age differences: the scores were higher for Despair, Adjustment, and Proactivity for females; for Aversion for males; and for Proactivity for students older than 23 years. The study suffers from limitations related to social desirability, gender imbalance, and self-selection effects in the recruitment.

Keywords: measures, coping strategies, pandemic, scale development, factor analysis, invariance analysis, Rasch model

OPEN ACCESS

Edited by:
Paola Gremigni, University of Bologna, Italy

Reviewed by:
Mel Win Khaw, Duke University, United States
Rudolf Debelak, University of Zurich, Switzerland

Correspondence:
Daniela Raccanello
daniela.raccanello@univr.it

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *15 June 2021*
Accepted: *10 August 2021*
Published: *09 September 2021*

Citation:
Burro R, Vicentini G, Rocca E, Barnaba V, Hall R and Raccanello D (2021) Development and Validation of the Robust - Pandemic Coping Scale (R-PCS). Front. Psychol. 12:725344. doi: 10.3389/fpsyg.2021.725344

INTRODUCTION

The outbreak of the COVID-19 pandemic posed a new and unexpected challenge worldwide, given the lack of readily available treatments or well-established solutions. The spread of the virus across countries and continents was marked by uncertainty and unpredictability (Vinkers et al., 2020), factors that are linked to increases in people's anxiety (Estes and Thompson, 2020). During the pandemic, policies for limiting the spread of the virus tended to be in a state of flux and were adjusted in response to new data and/or political pressure. The search for an effective vaccine was an ongoing priority. Because of the general uncertainty, there were major changes in the patterns of personal, social, cultural, and economic activities that had previously defined as "normal life." This was particularly so for university students given that they suffered very early and more extensively from the constraints aimed at preventing the spread of the virus. In many cases, they were forced to live within their dwellings or at some distance from family or friends; they had to respect social distancing and give up on-site university life and abruptly adapt to e-learning modalities. They were especially susceptible to experiencing economic uncertainty about both their current studies and their future profession. Given the key role of coping strategies for limiting and mitigating the negative consequences of disasters including pandemics, it is important to be able to separate the strategies that are most effective from those that are not so that people can be given the most effective psychological advice. To this end, we developed a self-report scale assessing a variety of pandemic-related coping strategies, using a dual approach combining the strengths of factor analyses and the Rasch model (Bond and Fox, 2007), i.e., the Robust - Pandemic Coping Scale (R-PCS).

Psychological Functioning, Coping Strategies, and Pandemics

Several studies, with samples ranging from younger to older adults, have shown that the COVID-19 pandemic has psychopathological consequences including anxiety, depression, panic behaviors, emotional distress, and various symptomatic responses (Arslan et al., 2020; Fitzpatrick et al., 2020; Gallagher et al., 2020; García-Portilla et al., 2020; Nicomedes and Avila, 2020; Rodríguez-Rey et al., 2020; Dozois and Mental Health Research Canada, 2021). Notably, it has been observed that the level of anxiety and fear was higher in the regions with higher COVID-19 rates (Fitzpatrick et al., 2020). Some authors have validated specific scales to assess fear and anxiety, such as the Fear of COVID-19 Scale (Ahorsu et al., 2020), which is a unidimensional scale based on seven items. This scale has already been employed in Bangladesh, Iran, Israel, Italy, Japan, New Zealand, Russia and Belarus, Saudi Arabia, Turkey, and Vietnam (Alyami et al., 2020; Bitan et al., 2020; Haktanir et al., 2020; Reznik et al., 2020; Sakib et al., 2020; Soraci et al., 2020; Wakashima et al., 2020; Winter et al., 2020). Using this scale, Satici et al. (2020) found that fear mediated the relation between intolerance of uncertainty and mental well-being: People who were more intolerant of uncertainties were also more afraid, and, in turn, their mental health was worse. Lee (2020) developed a 5-item scale using a sample of North American adults to assess anxiety – the Coronavirus and Anxiety Scale. This scale distinguishes between cognitive, behavioral, and psychological disturbances. Fear and anxiety can be considered as emotional reactions, which can, in fact, facilitate adjustment (Porcelli, 2020); however, extreme levels of fear can result in highly dysfunctional reactions and behaviors (Cheng and Tang, 2004). Some studies focused on university students and revealed increases in anxiety, depression, stress, post-traumatic stress disorder, perceptions of loneliness, fear, and worry (Cao et al., 2020; Elmer et al., 2020; Husky et al., 2020; Odriozola-González et al., 2020; Son et al., 2020; Tang et al., 2020).

Some of the negative consequences of pandemics can be mitigated by coping strategies. Coping is a multi-dimensional process for facing stressful situations (Lazarus and Folkman, 1984; Skinner et al., 2003; Skinner and Zimmer-Gembeck, 2007). According to Lazarus and Folkman's transactional model, there are three stages of evaluation that help individuals to adjust to threatening external situations. The first stage involves assessing the likely extent of damage or loss that might be incurred; the second stage involves identifying relevant coping strategies based on the individual and social resources that are available to the individual; and the third stage involves estimating the efficacy of each of these strategies. Coping strategies are commonly characterized as fitting into a typology based on the extent to which they focus on problems or emotions (Lazarus and Folkman, 1984) and the "fight or flight" responses they elicit (Schaefer and Moos, 1992). In the literature, there are several taxonomies classifying coping. In the attempt to propose a developmental classification including a wide range of coping strategies, Zimmer-Gembeck and Skinner (2011) elaborated a taxonomy incorporating three categories focused on three basic human needs, i.e., the needs for competence, relatedness, and autonomy (Deci and Ryan, 1985), that can be challenged or threatened by uncertain events. For each category, there are two adaptive and two maladaptive families of coping strategies that are typically activated in the face of perceived challenges or threats. Concerning the need for competence, the adaptive responses include problem solving (i.e., focusing on the problems with an analytical approach) and information seeking (i.e., searching for information by oneself or with others) while the maladaptive responses include helplessness (i.e., adopting a helpless or confused attitude when faced with situational demands) and escape (i.e., attempting to avoid the problem behaviorally or psychologically). In respect to the need for relatedness, the taxonomy includes self-reliance (i.e., focusing on emotional expression, understanding, or regulation) and support seeking (i.e., searching for concrete or psychological social support) in contrast to delegation (i.e., feeling of being out of control) and social isolation (i.e., withdrawing or refusing social contact). As for the need for autonomy, the taxonomy includes accommodation (i.e., attempting to adjust actively or by cognitively restructuring the situation) and negotiation (i.e., trying to increase the available options by contracting, persuading, or identifying priorities) in contrast with submission (i.e., adopting a passive attitude with intrusive thoughts or rumination) and opposition (i.e., demonstrating a marked refusal to cooperate). In **Table 1**, we report some examples of studies of coping strategies that can be coded according to this classification.

Some studies have recently examined how adults coped during the pandemics, exploring the links between coping strategies and some indicators of positive psychological functioning and/or psychopathological symptoms (Fullana et al., 2020; Wakashima et al., 2020; Bakker and van Wingerden, 2021; Park et al., 2021; Shamblaw et al., 2021). Shamblaw et al. (2021) explored adaptive strategies in a sample of American adults and found that a variety of strategies, such as active coping, positive reframing, planning, acceptance, emotional support, and the use of informational support was associated with lower levels of depression and higher levels of quality of life. Of the strategies they examined, they found that the most beneficial was positive reframing. In another study involving Americans, skills for regulating emotions, active coping, and distraction were significant predictors of a lower level of distress; while social support seemed less effective (Park et al., 2021). Other studies have focused on correlates of behavioral coping. Wakashima et al. (2020) found, among a Japanese sample, that the level of COVID-19-related fear was positively linked to protective behaviors (implying adherence to safety measures), stockpiling supplies, and monitoring one's health. Fullana et al. (2020) found that among a sample of Spanish adults, behavioral coping, such as having a healthy/balanced diet and not reading news/updates about COVID-19 very often was negatively associated with symptoms like anxiety and depression; in addition, following a routine, cultivating hobbies, and staying outdoors or looking outside were negatively linked to depression. All the strategies examined by these authors can be classified as challenge-related coping strategies focused on the need for competence, relatedness, and/or autonomy according to Zimmer-Gembeck and Skinner's (2011) classification.

Some data also suggest that most of the maladaptive coping strategies classified by Zimmer-Gembeck and Skinner (2011) for dealing with threats are, in fact, detrimental. In the research by Shamblaw et al. (2021), using avoidant coping (including strategies such as denial, substance use, venting, behavioral disengagement, distraction, and self-blame) was related to increases in depression and anxiety and decreases in quality of life. In a sample of Dutch adults, rumination was associated with lower well-being, operationalized in terms of depression, exhaustion, and less vigor (Bakker and van Wingerden, 2021). Research using a sample of Turkish participants found that rumination mediated the relation between intolerance of uncertainty and mental well-being, i.e., more intolerant people ruminated more, and in turn people who ruminated more had lower indicators of mental health; moreover, rumination was associated with increases in fear (Satici et al., 2020).

While most studies have used adult samples from a range of population groups, relatively little research has focused specifically on how university students coped with the COVID-19 pandemic (for exceptions, see Son et al., 2020;

TABLE 1 | Examples of coping strategies from previous studies and dimensions of the Robust - Pandemic Coping Scale (R-PCS) relative to each needs-based category.

	Need for competence		**Need for relatedness**		**Need for autonomy**	
	Examples of studies	**Dimensions of the R-PCS**	**Examples of studies**	**Dimensions of the R-PCS**	**Examples of studies**	**Dimensions of the R-PCS**
Challenges	Fullana et al., 2020: having a healthy/balanced diet, not reading COVID-19-related news very often. Park et al., 2021: active coping. Shamblaw et al., 2021: active coping, planning, informational support. Wakashima et al., 2020: protective behaviors, stockpiling supplies, monitoring health. Waselewski et al., 2020: prevention behaviors	Proactivity	Cao et al., 2020: social support. Park et al., 2021: regulation of emotions, social support. Shamblaw et al., 2021: emotional support. Son et al., 2020: support seeking. Waselewski et al., 2020: staying connected to people	Adjustment	Fullana et al., 2020: following a routine, cultivating hobbies, staying outdoors. Shamblaw et al., 2021: positive reframing, acceptance. Park et al., 2021: distraction. Son et al., 2020: sleeping longer, distracting by doing other tasks, meditation and breathing exercises, spiritual measures, keeping to routines, positive reframing. Waselewski et al., 2020: relaxing, thinking positively, keeping busy, following routines, cultivating hobbies, studying, having physical exercise	Adjustment
Threats	Shamblaw et al., 2021: avoidant coping. Son et al., 2020: ignoring the news	Despair	Elmer et al., 2020: isolation from social networks, lack of interaction, lack of emotional support, physical isolation. Panayiotou et al., 2021: difficulties in expressing and verbalizing emotions	Despair	Bakker and van Wingerden, 2021: rumination. Satici et al., 2020: rumination. Shamblaw et al., 2021: rumination. Son et al., 2020: drinking, smoking	Despair Aversion

Waselewski et al., 2020; Panayiotou et al., 2021). Among Cypriot university students, difficulties in expressing and verbalizing emotions on the one hand and difficulties in having access to a repertoire of emotion regulation strategies on the other hand predicted the decrease in quality of life due to the outbreak of the pandemic (Panayiotou et al., 2021). Waselewski et al. (2020) used a qualitative approach to explore how 14-to-24-year olds coped with the pandemic. Using content analysis, they identified a variety of strategies such as following prevention behaviors, staying connected to people, relaxing and thinking positively, keeping busy, following routines, cultivating hobbies, studying, or having physical exercise. American university students adopted support seeking and other strategies like ignoring the news, sleeping longer, distracting themselves by doing other tasks, drinking, or smoking, meditation and breathing exercises, spiritual measures, keeping to routines, and positive reframing strategies to cope with stress and anxiety due to COVID-19 (Son et al., 2020). Research among Swiss university students found that isolation from social networks, lack of interaction and emotional support, and physical isolation were associated with negative mental health (Elmer et al., 2020), while for Chinese university students, social support was negatively correlated with anxiety (Cao et al., 2020). However, notwithstanding the interest in coping strategies, there are, to our knowledge, no instruments specifically designed for use among university students to assess a wide range of coping strategies related to pandemics and/or epidemics.

Measurement of Disaster-Related Coping Strategies

Coping strategies are typically measured through self-report instruments, which have both disadvantages and advantages (Pekrun and Bühner, 2014). One limitation is that they only capture conscious psychological processes (although, coping strategies are usually conscious; Lazarus and Folkman, 1984; Compas et al., 2001). They may also be prone to desirability biases. However, being relatively cheap to implement and easy to be adapted to many different contexts, self-report instruments are still the most commonly used tools for accessing individuals' inner worlds.

Most of the published studies for assessing coping strategies related to COVID-19 used self-report questionnaires (e.g., Fullana et al., 2020; Wakashima et al., 2020; Bakker and van Wingerden, 2021; Park et al., 2021; Shamblaw et al., 2021) and only a few utilized open-ended questions with content analysis (e.g., Son et al., 2020; Waselewski et al., 2020). Generally, researchers who studied disaster-related coping strategies did not develop specific measures focused on disasters. As an exception, some authors validated the Power to Live with Disasters Questionnaire (PLDQ), a questionnaire measuring personality characteristics useful for coping with disasters, both in long (Sugiura et al., 2015) and short versions (Ishibashi et al., 2019). However, the PLDQ is not focused on disaster-related coping strategies. Developing a scale with pandemic-specific items was a response to the need for measuring coping strategies with an instrument relevant to the specific characteristics of this disaster. Pandemics and epidemics are typically characterized by a very long emergency phase, unlike, for example, earthquakes or tornados. During this phase, psychological interventions are urgently needed to assist people employ appropriate coping strategies. It follows, that it is important to have an instrument from which evidence-based recommendations can be made.

Despite the existence of some specific questionnaires to assess COVID-19-fear and anxiety (Ahorsu et al., 2020; Lee, 2020), to our knowledge, there is a lack of measures concerning coping strategies. Therefore, there was a need to develop a robust valid instrument focused on disaster-related coping strategies in general and on pandemic-related coping strategies in particular. By "robust and valid," we mean an instrument that meets the conditions of fundamental measurement, i.e., obtaining measurements not built from a foundation of other measurements, and which follows an additive logic (Bond and Fox, 2007). Fundamental measurement typically characterizes measurements made using basic units in the physical and natural sciences, while it is not so frequently encountered in the social sciences. To reach this objective, we used Rasch modelling (Rasch, 1960; Andrich, 1988) as the second stage of a two-stage approach to developing the scale in which exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were used to identify an initial set of items and then the items were fit to the Rasch model. Using such approach is particularly welcomed when the aim of the researchers is to diminish a large set of items to identify a lower number of scale scores (Christensen, 2021). In the literature, such approach has been amply used (e.g., Vidotto et al., 2010; Panella et al., 2012; Chiu et al., 2020). A major benefit of using Rasch scaling is that, when data can be shown to fit the Rasch model, it is possible to obtain for each item a score that is independent of characteristics of the sample of respondents and items, i.e., obtaining sample-free and test-free measures. These measures form an interval logit scale (Burro, 2016). There are several ordered steps that must be followed to apply the Rasch model and to verify its goodness of fit. First, one must evaluate the construct validity of the scale (Bretagnolle, 2002; Kang et al., 2018), the local independence (Marais, 2013; Debelak and Koller, 2020), the unidimensionality (Christensen et al., 2002; Smith, 2002), and the absence of differential item functioning (DIF; Tennant et al., 2004; Hagquist and Andrich, 2017), i.e., see that the instrument functions in the same way across different groups of participants. Second, one can use indices, such as the person separation index (PSI) or Cronbach's alpha (Wright, 2001; Kreiner and Christensen, 2013), to establish the reliability of the scale. Finally, one looks at the level of correspondence between the distribution of the calibrated items and that of the participants (Wright and Masters, 1982).

In some cases, it is possible that one or more of the assumptions concerning internal construct validity are not met. When this happens, one can implement an iterative procedure using different modification strategies (Linacre, 2002; Tennant and Conaghan, 2007), to account for violations of monotonicity (e.g., item rescoring), for violations of local independence (e.g., item grouping or "testlets" creation), and for the presence of DIF (e.g., item splitting). If the previous strategies do not work, one can delete critical items, and repeat the steps in the process. When all the assumptions are satisfied, the final step is to verify the fit of the model.

Aims

This study aimed at developing and testing the psychometric properties of a new scale, the R-PCS, designed to measure a range of coping strategies related to epidemics and pandemics. The scale was inspired by the classification proposed by Zimmer-Gembeck and Skinner (2011) that incorporates adaptive and maladaptive coping strategies pertaining to three functions, i.e., competence, relatedness, and autonomy.

The questionnaire contained three items for each of the 12 families of coping strategies proposed by Zimmer-Gembeck and Skinner (2011). We expected to identify different dimensions pertaining to categories reflecting adaptive or maladaptive coping strategies, that is, we expected to find at least one dimension focused on challenges and at least one dimension focused on threats. We also anticipated finding further dimensions linked to specific coping strategies (Hypothesis 1a). Moreover, we hypothesized that the factorial structure of the scale was invariant across gender and age of respondents (Hypothesis 1b). Then, we transformed each identified dimension applying the Rasch analysis.

The second aim was to study the validity of the R-PCS. Concerning the discriminant validity, we expected its dimensions to be independent (Hypothesis 2a). As regards the criterion validity, we expected that the dimensions would correlate with the factors of a scale designed to measure personality characteristics useful for coping with disasters, the PLDQ (Hypothesis 2b; Ishibashi et al., 2019). Pertaining to the predictive validity, we expected that the dimensions reflecting adaptive coping strategies would be positively related to enjoyment and negatively related to anger (Raccanello et al., 2021b) and vice versa for the dimensions reflecting maladaptive coping strategies (Hypothesis 2c) measured 2 months after the administration of the R-PCS.

The third aim was to examine interindividual differences in the R-PCS, examining possible differences related to gender and age of respondents.

MATERIALS AND METHODS

Participants

The sample consisted of 2,987 university students (M_{age} = 25.51 years, SD = 6.62; 79% females), from the University of Verona in Northern Italy. The participants were attending bachelor's degree courses (58%), master's degree courses (36%), or PhD and other specialization courses (6%). The sample was divided into two groups by splitting the total sample at the median age (24 years), with one group of 1,476 students (49% of the sample) being 23 years of age or younger and the other group of 1,511 students (51% of the sample) being older than 23 years. Concerning their health status as related to the 2020 COVID-19 pandemic at the time of the survey, 0.36% of them had been tested for novel coronavirus and had resulted positive, 1.43% had been tested for novel coronavirus and had resulted negative, and 98.21% had not been tested.

Procedure

The study was approved by the Director of the Head Office General Management of the University of Verona and by the Ethical Committee of the Department of Human Sciences of the same university (protocol n. 118846/2020). We sent an email to all the students attending the University of Verona during the academic year 2019–2020 (more than 25,000 students), inviting them to participate in an online survey on COVID-19 and emotions. The students gave their informed consent before participating. This work is part of a longitudinal study, in which we have administered the survey every 2 months since the outbreak of the pandemic. The first administration of the Italian language survey was between March 23 and April 1, 2020. In this paper, we also report data from a sub-sample of 998 students who participated in the second administration of the survey between May 18 and May 24, 2020.

Measures

Robust - Pandemic Coping Scale

We developed the R-PCS as follows. We conducted an extensive review of the literature on coping strategies used to deal with natural disasters. Two studies were particularly important in informing the development of the scale. One was a meta-analysis of relevant studies involving children and adolescents (Raccanello et al., 2019) and the other, a study in which adults reported adaptive strategies used to cope with earthquakes (Raccanello et al., 2021a). The literature review was followed by a process, in which four experts in developmental and educational psychology independently created a set of adaptive and maladaptive strategies that could be used to cope with the negative psychological consequences of pandemics, basing their work on previous research (e.g., Raccanello et al., 2019, 2020a,b, 2021a; Vicentini et al., 2020). A panel of judges (consisting of the four previously mentioned experts and two other experts in general psychology and education) discussed the set of items and retained 36 of the initial pool. These 36 items included adaptive or maladaptive coping strategies as identified by Zimmer-Gembeck and Skinner (2011). Some examples of the coping strategies to which the 36 items referred were as follows: adaptive strategies related to competence included problem solving, e.g., *Behaving in safe ways (for example washing my hands frequently)*, and information seeking/giving, e.g., *Looking for information from reliable sources* while among maladaptive strategies there was helplessness, e.g., *Thinking that nobody can help me*, and escape, e.g., *Pretending that there is no emergency*. Among the strategies focused on relatedness, adaptive strategies comprised, for example, self-reliance, e.g., *Keeping calm*, and support seeking/giving, e.g., *Collaborating with others* while among maladaptive strategies there was delegation, e.g., *Panicking*, and social isolation, e.g., *Being selfish*. Finally, adaptive coping strategies focused on autonomy included accommodation, e.g., *Keeping myself busy (for example playing or studying)*, and negotiation, e.g., *Creating new routines if usual ones cannot be followed*. Maladaptive strategies included submission, e.g., *Thinking that safety measures are not useful*, and opposition, e.g., *Thinking that media and politicians are exaggerating the situation*.

Each item was rated on a 5-point scale (1 = *never* and 5 = *always*) to indicate the frequency with which that strategy was used (*Think to how you have coped with emotions such as fear, sadness, and anger, that you could have felt since the outbreak of the pandemic. Please indicate how frequently you have used the following strategies*). In **Table 2**, we list the 20 items that were retained for the final version of the R-PCS after the statistical analyses.

Power to Live With Disasters Questionnaire

The participants completed the PLDQ (Ishibashi et al., 2019) that includes 16 items to be rated on a 5-point scale (1 = *not at all* and 5 = *very much*). The questionnaire measures eight personality characteristics useful for coping with disasters with two items for each factor: Leadership (e.g., *In everyday life, I often take the initiative to gather people together*), Problem solving (e.g., *When I am fretting about what I should do, I compare several alternative actions*), Altruism (e.g., *When I see someone having trouble, I have to help them*), Stubbornness (e.g., *I am stubborn and always get my own way*), Etiquette (e.g., *When someone has helped me or been kind to me, I clearly convey my feelings of gratitude*), Emotional regulation (e.g., *During difficult times, I endeavor not to brood*), Self-transcendence (e.g., *I am aware that I am alive, and have a sense of responsibility in living*), and Active well-being (e.g., *In everyday life, I have habitual practices that are essential for relieving stress or giving me a change of pace*). The Italian version was adapted through back-translation.

Achievement Emotions Adjective List

Two months after the administration of the R-PCS, the participants completed a brief version of the Achievement Emotions Adjective List (Raccanello et al., 2021b). The respondents rated the frequency with which they had felt enjoyment or anger in the previous 2 weeks using a 5-point scale (1 = *not at all* and 5 = *very much*).

Demographics

At the end of the questionnaire, we asked for the following demographic information: year of birth, gender, course (bachelor's degree courses, master's degree courses, PhD, and other specialization courses), and health status with respect to the 2020 COVID-19 pandemic (tested for novel coronavirus and positive, tested for novel coronavirus and negative, not tested). In the sample, 1,476 students were 23 years of age or younger, while 1,511 students were older than 23 years. For ethical reasons, these questions were not compulsory.

Data Analysis

The analyses were carried out using R software, Version 4.1.0 (R Core Team, 2021).

We conducted an EFA and a CFA, followed by Rasch analysis to assess the structure of the R-PCS (for a similar approach, see Raccanello et al., 2021c). We carried out the EFA on half of the sample and the CFA on the second half, after randomly splitting the initial sample into two sub-groups. To check whether the data were suitable for factor analyses, we ran the Bartlett's test of sphericity and the Kaiser-Meyer-Olkin test (KMO; *check_factorstructure*, function parameters R package; Lüdecke et al., 2020). The first determines whether there is a significant correlation in the data while the second examines the sample adequacy. Then, we ran a parallel analysis (Horn, 1965) and an optimal coordinates analysis (Ruscio and Roche, 2012) with the scree plot (*nScree* function, nFactors R package; Raiche and Magis, 2020), and a very simple structure analysis (*vss* function, psych R package; Revelle, 2021) to identify the appropriate number of factors for the EFA (Revelle and Rocklin, 1979). We applied both EFA (*fa* function, psych R package) and CFA (*cfa* function, lavaan R package; Rosseel, 2012) for ordinal data, beginning from a polychoric correlation matrix and using maximum likelihood and an oblique promax rotation for the EFA, and the DWLS estimator for the CFA. For the EFA, we used the root-mean-square error of approximation (RMSEA) and the comparative fit index (CFI), with RMSEA ≤ 0.08 and CFI ≥ 0.90 as threshold values to assess the goodness of fit; for the CFA, we also used the standardized root mean residual (SRMR) and the Tucker-Lewis index (TLI), respectively, with SRMR ≤ 0.08 and TLI ≥ 0.90 (Hu and Bentler, 1999; Marsh et al., 2005). Considering that when running a CFA, the minimum ratio between the number of observations and the number of parameters should be 5:1 or more, and preferably 10:1 (Kline, 2016), and, that in our case, we had 113 estimated parameters with 1,494 participants (about 13:1), the size of our sample was appropriate. We then conducted multigroup CFA by testing separate nested CFA models, analyzing: (1) the configural invariance model, allowing all the parameters to be freely estimated; (2) the metric invariance model, requiring invariant factor loadings; and (3) the scalar invariance model, additionally requiring invariant intercepts. To compare the models, we took into account differences in CFI, RMSEA, and SRMR: Support for invariance requires a change in CFI less or equal than 0.010, a change in RMSEA less or equal than 0.015, and a change in SRMR less or equal to 0.030 for testing metric invariance and less or equal to 0.010 for testing scalar invariance (Chen, 2007).

We then verified whether the data from the whole sample fitted the Rasch model (Andrich, 1988; *PCM* function, eRm R package; Mair et al., 2021), for each dimension identified by the CFA. First, we reviewed monotonicity, to check whether the thresholds, i.e., the transition points between two different scores, were correctly ordered. To do this, we used person-item maps. Then we checked for the possible presence of local

TABLE 2 | Item description in the English and Italian versions and factor loadings for the four factors.

Factor name	Item number	Italian version	English translation	Loadings Factor 1	Loadings Factor 2	Loadings Factor 3	Loadings Factor 4
Despair	6	Farsi prendere dal panico	Panicking	**0.761**	0.065	−0.111	−0.146
	27 (reversed)	Mantenere la calma	Keeping calm	**0.705**	0.024	−0.288	−0.185
	7	Pensare solo all'emergenza	Overthinking about the emergency	**0.558**	−0.100	0.115	−0.057
	32	Pensare che la situazione non migliorerà mai	Thinking that things will never get better	**0.505**	−0.013	0.114	0.167
	10	Pensare che nessuno possa aiutarci	Thinking that nobody can help me	**0.497**	−0.093	0.107	0.162
Adjustment	26	Inventarsi routine nuove se non si possono seguire quelle abituali	Creating new routines if usual ones cannot be followed	0.017	**0.676**	−0.134	−0.003
	29	Approfittare dell'occasione per coltivare i propri hobby	Taking the opportunity to cultivate hobbies	−0.143	**0.576**	−0.076	0.067
	5	Impegnarsi in qualcosa per distrarsi (ad esempio giocare o studiare)	Keeping myself busy (for example playing or studying)	−0.109	**0.534**	−0.160	−0.128
	30	Collaborare con gli altri	Collaborating with others	0.042	**0.407**	0.174	0.054
	25	Concentrarsi sulle cose veramente importanti (ad esempio la famiglia)	Focusing on things that are really important (for example family)	0.146	**0.404**	0.109	−0.055
Proactivity	36	Dare informazioni corrette, chiare e comprensibili	Giving correct, clear, and comprehensible information	−0.111	−0.082	**0.689**	0.046
	4	Informarsi tramite fonti affidabili	Looking for information from reliable sources	−0.041	−0.183	**0.622**	−0.074
	21	Aiutare e tranquillizzare chi è vicino a me	Helping and reassuring those around me	−0.038	0.205	**0.449**	0.153
	20	Chiedere informazioni in caso di dubbi sui comportamenti da tenere	In case of doubts, asking for information on appropriate behaviors	0.073	0.224	**0.405**	−0.025
	33	Comportarsi in modo sicuro (ad esempio lavando spesso le mani)	Behaving in safe ways (for example washing my hands frequently)	0.117	0.085	**0.401**	−0.285
Aversion	31	Pensare che i media e i politici stanno ingigantendo la situazione	Thinking that media and politicians are exaggerating the situation	−0.021	0.057	0.094	**0.473**
	12 (reversed)	Ricordarsi che rispettare le regole protegge la salute di tutti	Remembering that following the rules protects everybody's health	0.001	−0.170	−0.193	**0.457**
	3	Pensare che le misure di sicurezza adottate siano inutili	Thinking that safety measures are not useful	0.113	−0.077	0.120	**0.428**
	22 (reversed)	Seguire le indicazioni degli esperti	Following advice from experts	0.004	−0.128	−0.321	**0.416**
	34	Ignorare le ordinanze del Ministero della Salute	Ignoring the regulations from the Ministry of Health	0.011	−0.043	−0.072	**0.403**

dependence between the responses to the different items for each dimension. As the next step, we conducted one CFA for each dimension of the R-PCS (for a total of four separate unidimensional CFA) to confirm their unidimensionality – one of the assumptions of the Rasch model, which must be verified; then, we calculated the standardized P-DIF statistic (Dorans and Kulick, 1986) to determine whether there was a DIF across gender and age of respondents. Moreover, we examined the reliability calculating the PSI. After all these preliminary checks, we tested the fit of the data to the Rasch model using Andersen's likelihood ratio test (Andersen, 1973). We then examined the item performance studying infit (i.e., mean square inlier-sensitive fit) and outfit (i.e., mean square outlier-sensitive fit), considering the thresholds for rating scale surveys (Wright and Linacre, 1994). When item mean-squares are higher than 1.40, it means that the items underfit the Rasch model; when the mean-squares are lower than 0.60, it means that the items overfit the Rasch model. At the end of this series of steps, we transformed the raw scores into logit scores for use in all the following analyses.

We then examined the discriminant validity of the R-PCS (second aim) following the approach of Rönkkö and Cho (2020). We calculated the intercorrelations and the descriptive statistics between its dimensions and the factors of the PLDQ to assess the discriminant validity of the R-PCS. Correlations between 0.10 and 0.30 can be considered as small, between 0.30 and 0.50 as moderate, and higher than 0.50 as large (Cohen, 1988). We checked the eight-factor structure of the scale through a CFA. Subsequently, we investigated whether the four dimensions of the R-PCS (logit scores) predicted 2-month-after enjoyment and anger ratings, running four linear mixed models (LMM; *lmer* function, lme4 R package;

Bates et al., 2015), with each separate dimension as numeric fixed effects, participants as the random effect, and emotions (enjoyment and anger) as dependent variables. Finally, we examined gender and age differences on the dimensions of the R-PCS, conducting a LMM, with gender (males and females), age (23 years of age or less and older than 23 years), and dimensions of the R-PCS as categorical fixed effects, and participants as the random effect. The logit scores of each dimension of the R-PCS were the dependent variables. We performed a type III analysis of variance table with Satterthwaite's method. We used the Bonferroni correction for *post-hoc* tests (*emmeans* function, emmeans R package; Lenth, 2021). The level of significance was $p < 0.05$.

RESULTS

Structure of the R-PCS

Exploratory and Confirmatory Factor Analyses

The preliminary analyses conducted on half sample ($n = 1,493$) indicated that the data were appropriate for factor analysis (Bartlett's test of sphericity: $X^2(630) = 11639.75$, $p < 0.001$; KMO = 0.88). The parallel analysis and the optimal coordinates analysis suggested that the most appropriate number of factors to extract was six (see the scree plot in **Figure 1A**), while the analysis of the very simple structure suggested that it was four (**Figure 1B**).

We ran the EFA, extracting six factors. The fit indexes revealed the adequacy of the model, CFI = 0.99, RMSEA = 0.02. We used items with factor loadings larger than 0.40 to define each factor. Given that the fifth factor had only two items with loadings larger than 0.40, and the sixth factor had only one item with a loading larger than 0.40, and that the very simple structure analysis had suggested extracting four factors, we decided to keep only the first four factors for subsequent analyses. The factor loadings of the selected items are shown in **Table 2**. Note that items 12, 22, and 27 were reverse scored before the EFA.

The first dimension (which explained the 26% of the variance) included one item on helplessness (item 10), two items on delegation (item 6 and 27 – the latter was developed to assess self-reliance, but given that the score was reversed, we can interpret it in terms of its opposite strategy, i.e., delegation), and two items on submission (items 7 and 32). This dimension, we called as "Despair," referred to the level of immobility of people who are overwhelmed by panic and lose any hope, both cognitively and emotionally. They are blocked and continue to ruminate on the emergency: They do not react, except through panic and despair, and they are convinced that there are no solutions to the problem. Overall, this factor comprises threats to the three basic needs, i.e., competence, relatedness, and autonomy.

The second dimension (which explained the 19% of the variance) included two items on accommodation (items 5 and 29) and two items on negotiation (items 25 and 26), all pertaining to challenges to the need for autonomy. In addition, it had one item related to support seeking/giving (item 30), which is also reflected in item 25 that mentions family relationships. This dimension, named as "Adjustment," refers to the level with which people react in an adaptive and constructive way within the broad context of both individuals and activities. Overall, this factor focused on the challenges to relatedness and autonomy.

The third dimension (which explained the 19% of the variance) included one item related to problem solving (item 33) and three items related to information seeking/giving (items 4, 20, and 36). These items all concern challenges to the need for competence. Also, one item originally developed for assessing support giving (item 21) loaded on this dimension, which nevertheless referred to performing active prosocial actions to solve a problem. We termed this dimension "Proactivity," reflecting the level to which people activate themselves to find solutions to problems, through behaviors aimed at protecting health, seeking and understanding reliable information, and helping others.

The fourth dimension (which accounted for the 17% of the variance) included two items pertaining to opposition (items 31 and 34) and other three items that had been developed, initially, for assessing problem solving (item 22, reversed), negotiation (item 12, reversed), and submission (item 3), respectively. All the items included reference to explicit opposition to rules. We called this dimension as "Aversion" as it reflected the extent to which people fail to accept the health protection rules established by competent authorities. This factor focused on the threats to the need for autonomy.

In **Table 1**, we show the four dimensions and their relationship to Zimmer-Gembeck and Skinner's (2011) classification; in particular, their potential links to what have been described as challenges/threats and the three basic needs. Following the EFA, we performed a CFA on the other half of the sample ($n = 1,494$) with four factors (**Figure 2**). The fit indexes, CFI = 0.950, TLI = 0.940, RMSEA = 0.062, and SRMR = 0.067, indicated that there was a relatively good fit between the hypothesized model and the observed data. Therefore, the CFA supported the idea that the R-PCS is characterized by four distinct dimensions, two related to challenges (i.e., Adjustment and Proactivity) and two related to threats (i.e., Despair and Aversion), confirming Hypothesis 1a.

Finally, to test MI, we conducted a sequence of gradually more restrictive tests to verify the configural, metric, and scalar invariance (**Table 3**). The results indicated that for the R-PCS the hypothesized measurement model was invariant and generalizable across both gender and age of respondents, for all the levels of invariance, corroborating Hypothesis 1b.

Rasch Analysis

We applied the Rasch analysis using the partial credit model (Masters and Wright, 1997), separately for each dimension of the R-PCS, i.e., Despair, Adjustment, Proactivity, and Aversion. First, we checked the monotonicity using person-item maps (**Figure 3**). A person-item map represents the relation between the location of a person's coping strategies and the items' discriminatory capacities. In **Figure 3**, the parameter related to the person (i.e., coping strategies) varies from lower scores on

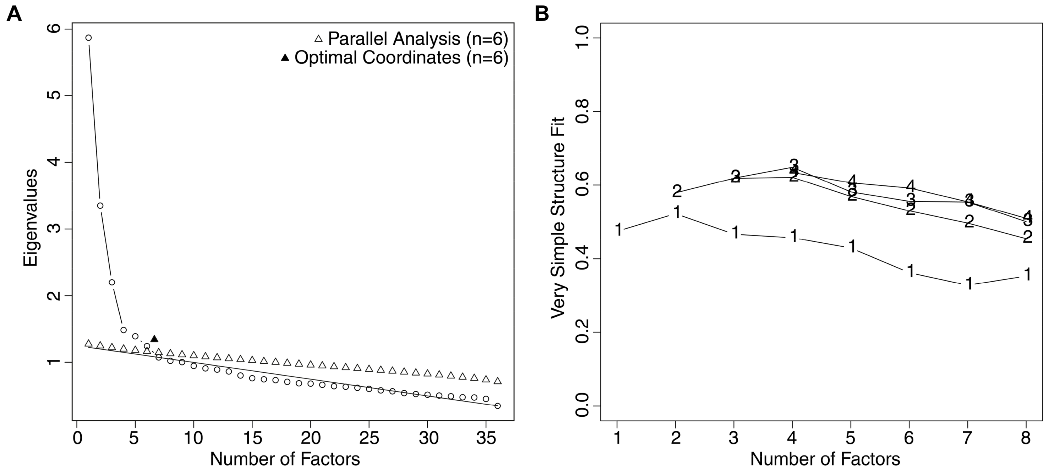

FIGURE 1 | (A) Parallel analysis scree plot and optimal coordinates plot related to the R-PCS, suggesting that the best solution was at six factors. **(B)** The very simple structure plot, indicating that the best solution was at four factors.

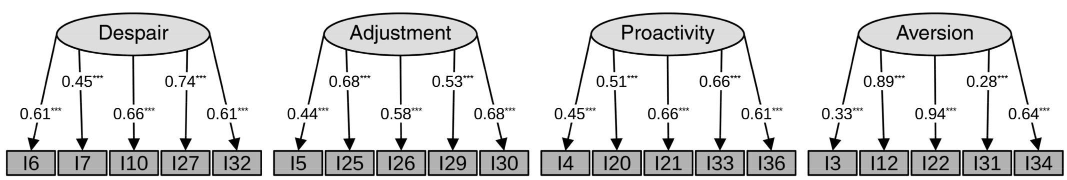

FIGURE 2 | Factorial model of the R-PCS. The digits represent standardized factor loadings. ***$p < 0.001$.

TABLE 3 | Results of invariance analyses across gender (males, females) and age (younger than or as old as 23 years, older than 23 years).

Groups	Model	CFI	RMSEA	SRMR	Δ CFI	Δ RMSEA	Δ SRMR
Gender	Configural invariance	0.947	0.062	0.066	0.003	<0.001	<0.001
	Metric invariance	0.945	0.061	0.068	0.002	<0.001	0.001
	Scalar invariance	0.941	0.059	0.067	0.004	<0.002	<0.001
Grade	Configural invariance	0.947	0.063	0.068	0.003	0.001	<0.001
	Metric invariance	0.946	0.062	0.068	<0.001	0.001	<0.001
	Scalar invariance	0.945	0.058	0.068	0.001	0.004	<0.001

N = 1,494. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; Δ CFI/RMSEA/SRMR = change in CFI/RMSEA/SRMR.

the left to higher scores on the right of the figures. The maps indicated that the scores of four items in the Adjustment (item 5), Proactivity (item 36), and Aversion (items 12 and 22) dimension did not have ordered thresholds (**Figures 3B,D,F**). Thus, we rescored them because they violated the monotonicity assumptions (we specify that we had reversed the scores of items 12 and 22 before recoding them). For items 5 and 36, the response scale changed from 1, 2, 3, 4, 5 to 1, 1, 2, 3, 4; for items 12 and 22, it changed from 1, 2, 3, 4, 5 to 1, 2, 3, 4, 4. This resulted in the items in the three factors, i.e., Adjustment, Proactivity, and Aversion, meeting the monotonicity requirements (**Figures 3C,E,G**). We then examined the correlations between the item residuals (Despair: $-0.36 < r < 0.10$; Adjustment: $-0.39 < r < -0.14$; Proactivity: $-0.43 < r < -0.01$; Aversion: $-0.45 < r < 0.12$), which were never larger than 0.30; therefore, there was no evidence of local dependence. The CFA conducted separately for each of the four dimensions indicated that each of them was unidimensional (Despair: CFI = 0.997, TLI = 0.989, RMSEA = 0.057, and SRMR = 0.026; Adjustment: CFI = 0.981, TLI = 0.953, RMSEA = 0.086, and SRMR = 0.044; Proactivity: CFI = 0.998, TLI = 0.994, RMSEA = 0.029, and SRMR = 0.016; Aversion: CFI = 0.981, TLI = 0.952, RMSEA = 0.076, and SRMR = 0.051). Then, we calculated the standardized P-DIF statistic separately for gender (Despair: $-0.071 < \mathrm{DIF} < 0.086$; Adjustment: $-0.026 < \mathrm{DIF} < 0.031$; Proactivity: $-0.013 < \mathrm{DIF} < 0.001$; Aversion: $-0.079 < \mathrm{DIF} < 0.049$) and age (Despair: $-0.016 < \mathrm{DIF} < 0.016$; Adjustment: $-0.009 < \mathrm{DIF} < 0.003$; Proactivity: $-0.004 < \mathrm{DIF} < 0.003$; Aversion: $-0.015 < \mathrm{DIF} < 0.032$). Given that

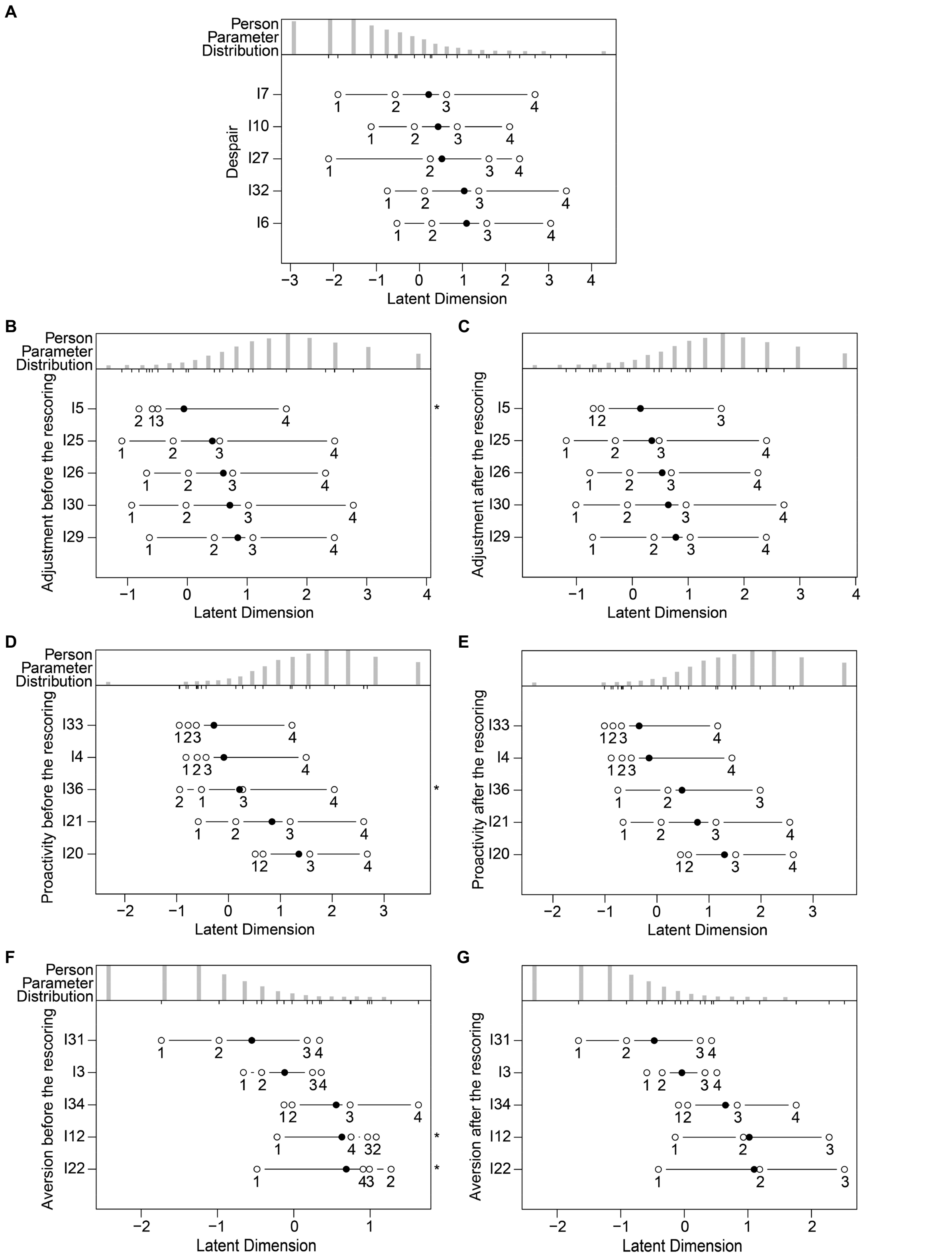

FIGURE 3 | Person-item maps relating to the five items of each dimension of the R-PCS for: **(A)** Despair, without rescoring; **(B)** Adjustment, before rescoring; **(C)** Adjustment, after rescoring; **(D)** Proactivity, before rescoring; **(E)** Proactivity, after rescoring; **(F)** Aversion, before rescoring; **(G)** Aversion, after rescoring. We represented the locations of the items' discriminatory capacities through solid circles and the thresholds through open circles. We indicated the items with non-ordered thresholds with asterisks.

they fell between −0.10 and 0.10, we can say that each dimension functioned similarly for males and females and for younger and older participants. Concerning reliability, all the PSI were adequate (Despair: 0.71; Adjustment: 0.71; Proactivity: 0.70; Aversion: 0.70).

At this point, we carried out four separate Andersen's likelihood ratio tests (Despair: $X^2(19) = 23.000$, $p = 0.237$; Adjustment: $X^2(18) = 12.800$, $p = 0.802$; Proactivity: $X^2(18) = 21.800$, $p = 0.242$; Aversion: $X^2(17) = 12.394$, $p = 0.776$), which showed that for each dimension the data fit the Rasch model. Also, the infit and outfit mean square statistics for each item of each dimension (**Table 4**) confirmed that the data were predicted by the model (all the values fell between 0.60 and 1.40). Subsequently, for each dimension and for each participant, we summed the raw scores of the five items, and we obtained a global score. Finally, we transformed the raw scores into an interval logit scale (Masters and Wright, 1997), as shown in the conversion table (**Table 5**). To increase the usability of the scale, the logit scores were scaled from 1 to 10 (considering that four items, i.e., 5, 12, 22, and 36, were rescored).

The intercorrelations and the descriptive statistics of the four dimensions of the R-PCS are shown in **Table 6**. The McDonald's omega reliability indexes were 0.79, 0.72, 0.73, and 0.70, respectively, for Despair, Adjustment, Proactivity, and Aversion, suggesting that the scale had an acceptable reliability. In **Figure 4**, we plot the intercorrelations between the four dimensions of the R-PCS (logit scores), showing that the four-factor solution is characterized in terms of two categories, adaptive and maladaptive – i.e., the two adaptive dimensions (Adjustment and Proactivity) correlated negatively with the two maladaptive dimensions (Despair and Aversion).

TABLE 4 | Infit and outfit mean square statistics (MSQ) of each item of each dimension of the R-PCS.

Factor name	Item number	Infit-MSQ	Outfit-MSQ
Despair	6	0.67	0.61
	27 (reversed)	0.93	0.93
	7	0.80	0.82
	32	0.87	0.89
	10	0.90	0.90
Adjustment	26	0.95	0.95
	29	0.72	0.75
	5	0.79	0.80
	30	0.83	0.86
	25	0.84	0.84
Proactivity	36	0.84	0.90
	4	0.79	0.81
	21	0.80	0.80
	20	0.81	0.82
	33	0.84	0.88
Aversion	31	0.79	0.77
	12 (reversed)	0.82	0.79
	3	0.81	0.78
	22 (reversed)	0.87	0.86
	34	0.91	0.95

TABLE 5 | Conversion table from raw scores of the items of the R-PCS to logit scores, separately for each dimension.

	Logit scores			
Raw scores	Despair	Adjustment	Proactivity	Aversion
5	1.000	1.000	1.000	1.000
6	1.812	1.737	1.637	1.969
7	2.560	2.400	2.253	2.855
8	3.043	2.925	2.755	3.401
9	3.414	3.234	3.267	3.807
10	3.725	3.658	3.576	4.139
11	4.000	3.958	3.861	4.429
12	4.254	4.239	4.138	4.695
13	4.494	4.511	4.414	4.947
14	4.729	4.781	4.697	5.197
15	4.963	5.058	4.990	5.453
16	5.200	5.346	5.293	5.722
17	5.442	5.654	5.614	6.016
18	5.695	5.989	5.957	6.348
19	5.963	6.363	6.334	6.735
20	6.253	6.791	6.760	7.377
21	6.579	7.300	7.266	8.089
22	6.962	7.947	7.916	8.956
23	8.200	8.933	8.916	10.000
24	9.046	10.000	10.000	–
25	10.000	–	–	–

Validity of the R-PCS

Discriminant Validity

To investigate discriminant validity, we examined the latent correlations based on the results of the CFA and their confidence intervals (CI) using a significance level of 5%, since the CFA model considers the measurement error (**Table 3**). We took into account the CI_{CFA} checking if the upper limit of the CI for each latent correlation was lower than 0.80. In most of the cases the upper limits were below the cut-off, ranging from 0.27 to 0.80, confirming the discriminant validity for the measures (Rönkkö and Cho, 2020). There was a significant, positive, and moderate correlation between the two adaptive dimensions (Adjustment and Proactivity), and a positive and small correlation between the two maladaptive dimensions (Despair and Aversion). The adaptive dimensions negatively correlated with the maladaptive dimensions, and in most of the cases the correlations were small (**Table 6**). Therefore, our data revealed that the four dimensions were separable, confirming Hypothesis 2a.

Criterion Validity

We conducted a preliminary CFA to test the factorial structure of the PLDQ. The model with eight factors was adequate, CFI = 0.974, TLI = 0.958, RMSEA = 0.052, and SRMR = 0.052. We then calculated the intercorrelations between the four dimensions of the R-PCS and the eight factors of the PLDQ (**Table 6**).

As regards the two adaptive dimensions, there were significant, positive, and moderate correlations between Adjustment and Proactivity on the one hand and most of the factors of the PLDQ on the other hand, with some exceptions (i.e., the correlation between Adjustment and Stubbornness was not significant; the one between Proactivity and Stubbornness was small). Concerning the two maladaptive dimensions, there were

TABLE 6 | Intercorrelations and descriptive statistics (means, M; standard deviations, SD; 95% confidence intervals, CI) for the four dimensions of the R-PCS (logit scores) and the eight factors of the power to live with disasters questionnaire (PLDQ).

S.no.	Variable	1	2	3	4	5	6	7	8	9	10	11	12
1.	R-PCS – Despair	–											
2.	R-PCS – Adjustment	−0.22**	–										
3.	R-PCS – Proactivity	−0.15**	0.45**	–									
4.	R-PCS – Aversion	0.12**	−0.19**	−0.31**	–								
5.	PLDQ – Leadership	−0.08**	0.33**	0.32**	−0.13**	–							
6.	PLDQ – Problem solving	−0.18**	0.35**	0.35**	−0.13**	0.32**	–						
7.	PLDQ – Altruism	−0.03	0.31**	0.31**	−0.14**	0.40**	0.29**	–					
8.	PLDQ – Stubbornness	0.01	0.03	0.04*	0.12**	0.06*	0.11**	0.03	–				
9.	PLDQ – Etiquette	−0.14**	0.37**	0.34**	−0.15**	0.49**	0.28**	0.40**	−0.02	–			
10.	PLDQ – Emotional regulation	−0.37**	0.38**	0.30**	−0.12**	0.21**	0.38**	0.16**	−0.04*	0.26**	–		
11.	PLDQ – Self-transcendence	−0.12**	0.42**	0.39**	−0.22**	0.34**	0.34**	0.38**	−0.01	0.37**	0.33**	–	
12.	PLDQ – Active well-being	−0.13**	0.43**	0.32**	−0.05*	0.33**	0.42**	0.22**	0.08**	0.32**	0.35**	0.41**	–
	M	3.13	6.47	6.84	2.98	3.21	3.52	4.22	2.67	3.54	3.30	3.95	3.42
	SD	1.17	1.38	1.47	1.21	0.83	0.72	0.63	0.82	0.82	0.82	0.74	0.81
	95% CI	[3.09, 3.17]	[6.42, 6.52]	[6.79, 6.89]	[2.94, 3.03]	[3.21, 3.28]	[3.49, 3.54]	[4.19, 4.24]	[2.64, 2.70]	[3.51, 3.57]	[3.26, 3.32]	[3.92, 3.97]	[3.39, 3.45]

*$p<0.05$; **$p<0.01$; and **$p<0.001$.

significant, negative, and small correlations between Despair and Aversion on the one hand and most of the factors of the PLDQ on the other hand, with some exceptions. The correlations between Despair and Altruism, and between Despair and Stubbornness, were not significant; the one between Despair and Emotional regulation was moderate; and the one between Aversion and Stubbornness was positive. Hence, the data indicated the criterion validity of the R-PCS, supporting Hypothesis 2b.

Predictive Validity

Adjustment was related positively to enjoyment, $\beta = 0.049$, $p = 0.044$, and negatively to anger, $\beta = -0.050$, $p = 0.043$. Proactivity was positively related to enjoyment, $\beta = 0.044$, $p = 0.050$. Both Despair, $\beta = 0.150$, $p < 0.001$, and Aversion, $\beta = 0.124$, $p < 0.001$, were positively related to anger. Therefore, our data supported Hypothesis 2c.

Gender and Age Differences

The LMM revealed a significant effect of the R-PCS dimensions, $F(3, 11{,}948) = 4718.478$, $p < 0.001$, $\eta^2_p = 0.83$. The *post-hoc* tests indicated that the scores were higher for Adjustment ($M = 6.47$, $SD = 1.38$, 95% CI [6.42, 6.52]) and Proactivity ($M = 6.84$, $SD = 1.47$, 95% CI [6.79, 6.89]) compared to Despair ($M = 3.13$, $SD = 1.17$, 95% CI [3.09, 3.17]; Adjustment vs. Despair, $z = 80.15$, $p < 0.001$; Proactivity vs. Despair, $z = 90.12$, $p < 0.001$) and Aversion ($M = 2.98$, $SD = 1.21$, 95% CI [2.94, 3.03]; Adjustment vs. Aversion, $z = 77.51$, $p < 0.001$; Proactivity vs. Aversion, $z = 87.48$, $p < 0.001$).

Also gender, $F(1, 11{,}948) = 87.303$, $p < 0.001$, $\eta^2_p = 0.03$, and age, $F(1, 11{,}948) = 6.770$, $p = 0.009$, $\eta^2_p = 0.01$, had significant effects, in turn moderated by two significant two-way interactions, gender X R-PCS dimensions, $F(3, 11{,}948) = 51.616$, $p < 0.001$, $\eta^2_p = 0.05$ (**Figure 5A**), and age X R-PCS dimensions, $F(3, 11{,}948) = 3.841$, $p = 0.009$, $\eta^2_p = 0.01$ (**Figure 5B**). Examining the *post-hoc* tests, we found that males had lower scores than females for Despair (males: $M = 2.72$, $SD = 1.12$, 95% CI [2.63, 2.80]; females: $M = 3.24$, $SD = 1.16$, 95% CI [3.19, 3.29]; $z = -9.00$, $p < 0.001$), Adjustment (males: $M = 6.04$, $SD = 1.34$, 95% CI [5.94, 6.15]; females: $M = 6.58$, $SD = 1.37$, 95% CI [6.52, 6.63]; $z = -9.07$, $p < 0.001$), and Proactivity (males: $M = 6.54$, $SD = 1.45$, 95% CI [6.43, 6.66]; females: $M = 6.92$, $SD = 1.47$, 95% CI [6.86, 6.98]; $z = -6.58$, $p < 0.001$), while they had higher scores for Aversion (males: $M = 3.26$, $SD = 1.20$, 95% CI [3.17, 3.36]; females: $M = 2.91$, $SD = 1.20$, 95% CI [2.86, 2.96]; $z = 5.96$, $p < 0.001$). Concerning age, only for Proactivity the scores were lower, $z = -4.19$, $p < 0.001$, for younger ($M = 6.73$, $SD = 1.48$, 95% CI [6.65, 6.80]) compared to older students ($M = 6.95$, $SD = 1.46$, 95% CI [6.88, 7.02]).

DISCUSSION

We developed a brief, reliable, and valid scale to assess adaptive and maladaptive coping strategies related to pandemics and epidemics, based on Zimmer-Gembeck and Skinner's (2011)

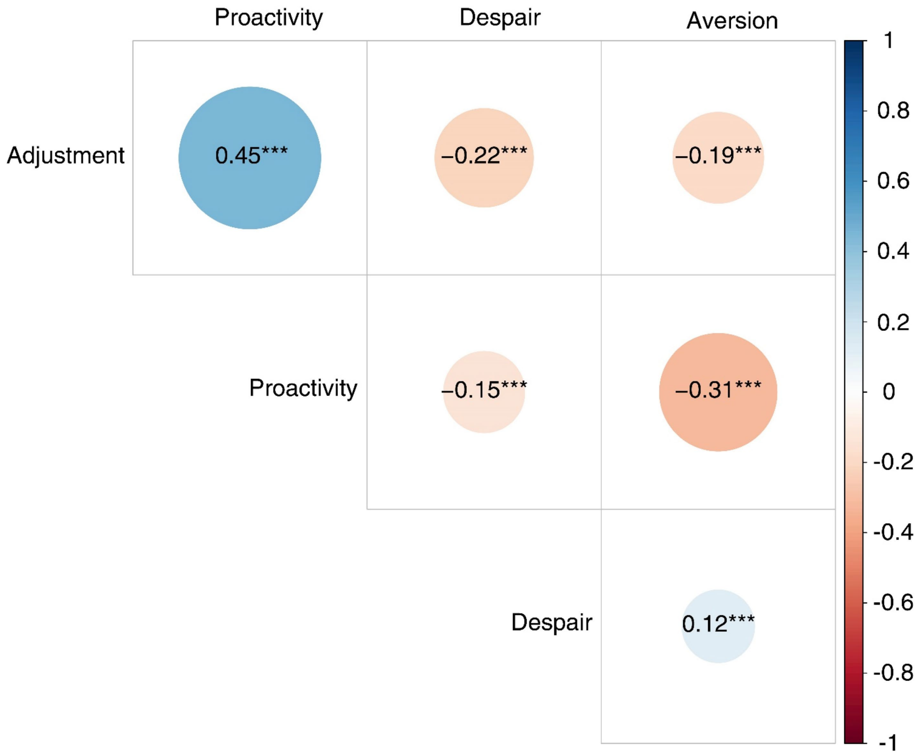

FIGURE 4 | Plot of the intercorrelations between the four dimensions of the R-PCS (logit scores).

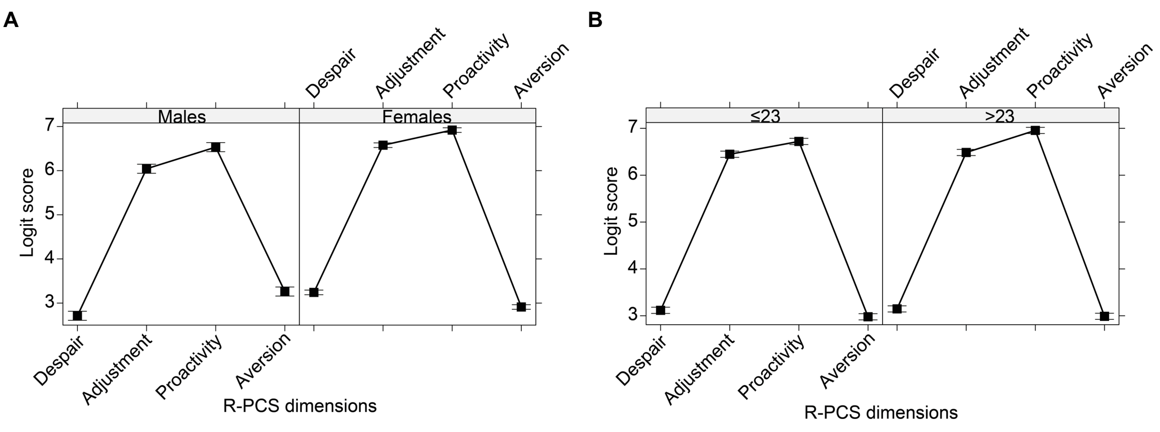

FIGURE 5 | R-PCS logit scores of the four dimensions, according to **(A)** gender and **(B)** age of respondents. The bars represent the 95% CI.

classification. Importantly, by fitting a Rasch model to the data collected for this project, we have created a scale in which the item scale values are independent of the sample of people who completed the items during scale development. This means that the items will have the same scale properties in any sample in which it is used and an asset for researchers working in the area.

The relevance of this paper can be appreciated considering methodological (first and second aim), theoretical (third aim), and applied perspectives.

Our first aim was to examine the structure of the R-PCS. Through a dual approach combining an EFA and a CFA, we identified the factorial structure of the R-PCS, including four first-order dimensions, namely Despair, Adjustment, Proactivity, and Aversion. Confirming Hypothesis 1a, we found two dimensions focused on challenges and two dimensions focused on threats. Based on the psychological literature (Zimmer-Gembeck and Skinner, 2011), we could speculate that the first two were adaptive while the other two were maladaptive. Adjustment included items pertaining to challenges to relatedness and autonomy, while Proactivity items concerning challenges to competence. Despair comprised items referred to threats covering all the three needs, i.e., competence, relatedness, and autonomy, while Aversion was specifically focused on items on threats to individuals' autonomy. We could consider that the overall capacity to overcome stressful events can result from the combination of the different dimensions. Therefore, future studies could explore individuals' profiles concerning how they react to traumatic events, such as the COVID-19 pandemic, and other similar stressful events. They could identify which combinations of different levels of endorsement of Despair, Adjustment, Proactivity, and Aversion are associated with individuals' emotional reactions or other indicators of mental disturbance or positive psychological functioning.

We also tested the measurement invariance of the R-PCS, which was invariant at the configural, metric, and scalar levels both across gender and across age. We then applied the Rasch model, transforming the scores of each dimension into interval level measures, with all the advantages related to the principles of the fundamental measurement (Rasch, 1960; Andrich, 1988; Bond and Fox, 2007; Burro, 2016).

Concerning the second aim, our findings showed the discriminant validity of the R-PCS, revealing that the four identified dimensions were independent and separable, confirming Hypothesis 2a. Moreover, the analysis of the correlations with the PLDQ showed good criterion validity, also supporting Hypothesis 2b. Concerning predictive validity, we examined the relationships between the four dimensions of the R-PCS and two emotions measured after 2 months, and our data confirmed Hypothesis 2c. Adjustment and Proactivity appeared adaptive, being both positively related to enjoyment while Adjustment was also negatively related to anger; Despair and Aversion seemed maladaptive, i.e., positively related to anger. Therefore, our findings supported the theoretical assumptions (Zimmer-Gembeck and Skinner, 2011) of the adaptive nature of coping strategies focused on challenges (Adjustment and Proactivity), and the maladaptive nature of those focused on threats (Despair and Aversion).

Among our sample of Italian university students (third aim), the scores of the two dimensions focused on challenges, i.e., Adjustment and Proactivity, were higher than the scores of the two dimensions focused on threats, i.e., Despair and Aversion. These findings suggest that, during the first wave of the COVID-19 pandemic, the students still perceived that they had adaptive resources, which enabled them to face such a stressful event. Future research could investigate how and whether these resources changed in the face of the continuous and persistent emergency phase of the pandemic. Moreover, females were characterized by higher scores on Despair, Adjustment, and Proactivity than males and vice versa for Aversion. The findings concerning Despair and Aversion are in line with the previous literature on the prevalence of internalizing behaviors for females and of externalizing behaviors for males, especially since adolescence (Rosenfield, 2000). Finally, Proactivity increased for older students. In any case, the effect sizes of all these differences were quite low. Therefore, these differences could be explored further, examining possible links with other constructs, and specifically other differential adjustments to the pandemic over time.

At the applied level, being able to measure pandemic-related coping strategies is of basic relevance for subsequent interventions. The literature has documented that, at least for university students, being able to verbalize emotions and having access to a range of coping strategies is positively linked to their quality of life during the pandemic (Panayiotou et al., 2021). Therefore, during and after disasters it is a priority to have instruments to detect how people are reacting and to identify in which areas they have difficulties. For example, the R-PCS could be used to monitor university students' coping strategies during the different phases of a pandemic to inform policy decisions. In addition, it could be applied before and after interventions aimed at supporting adults in coping with the emotional challenges of a pandemic, to assess the efficacy of the interventions. Moreover, it could be used with patients to help in prescribing appropriate individualized interventions aimed at fostering emotional competence.

This study suffers from some limitations. One limitation is that the final version of the R-PCS did not include all the coping strategies of Zimmer-Gembeck and Skinner's (2011) classification, because of the mediocre factorial loading of the corresponding items. Specifically, it did not comprise escape and social isolation. For both coping strategies, we could speculate that this could be linked to the contents of the items themselves. For example, the escape-related item *Pretending that there is no emergency* could refer more to a psychiatric symptom than to a proper coping strategy. In addition, the social-isolation item *Being selfish* could be particularly affected by social desirability biases and therefore being associated with a different pattern of responses compared to the other items. In future studies, we could reformulate the items relating to the excluded coping strategies to expand the scale. It is also worth noting that, in the psychological literature, there are several classifications of coping, and therefore we do not claim that our scale captures every type of coping. Moreover, the responses to the whole questionnaire could have suffered from a social desirability bias. One way to deal with this issue is to assure the confidential nature of the data, fostering people's tendency to trust the researchers (Pekrun and Bühner, 2014). Another limitation relates to the gender imbalance in our sample, with the majority of participants being female. It is worth noting that such imbalance is consistent with the percentages of females (63.9%) attending the University of Verona during the academic year 2019–2020. Moreover, the unbalanced composition of our sample could also be due to the fact that females were more prone to spend some time in an activity that was seen as having prosocial aims, i.e., completing a survey to increase knowledge on the emotional consequences of the pandemic. We could speculate that this is in line with gender stereotype according to which girls engage in more prosocial behaviors (Hastings et al., 2015). Finally, the recruitment of the sample could have been biased by self-selection effects, and we could not investigate the reasons underlying the decision not to respond to the survey; a critical aspect of most of the research on these topics conducted during the 2020 pandemic. In addition, we specify that the generality of our findings can be extended to students of the same age of similar socio-cultural contexts who are living the emergency phase of a disaster with characteristics similar to the ongoing pandemic. On the one hand, we unfortunately note that, currently, the COVID-19 pandemic is still on course, and therefore, there are potentially many students who are in a situation similar to the one that characterized our sample when they participated to our survey. Moreover, the R-PCS could be used also in post-pandemic assessment with samples with similar characteristics. On the other hand, future research could replicate our findings with samples varying for other characteristics, to favor the robust advancement of the scientific knowledge about how to cope in front of disasters. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context (Simons et al., 2017). Notwithstanding these limitations, the current study offers a new instrument to assess pandemic-related coping strategies, the R-PCS, whose psychometric properties benefit from the strengths of the Rasch model. Even if the scale has been developed during a disaster, such as the COVID-19 pandemic, it can be used to measure coping strategies in all the phases of the disaster management cycle, i.e., before, during, and after a pandemic or an epidemic. Always considering Zimmer-Gembeck and Skinner's (2011) classification as the theoretical framework, in the future the scale can also be adapted to other disasters and for different age groups.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Committee of the Department of Human Sciences of the University of Verona (Italy). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RB, DR, and GV contributed to conception, design of the study, and organized the database. RB performed the statistical analysis. RB and DR wrote the first draft of the manuscript. RB, DR, GV, VB, ER, and RH wrote sections of the manuscript. All authors contributed to manuscript revision, and read and approved the submitted version.

FUNDING

This work was supported by the Fondazione Cariverona, Bando Ricerca Scientifica 2017, Italy.

ACKNOWLEDGMENTS

The authors want to thank all the participants and the University of Verona for its support.

REFERENCES

Ahorsu, D. K., Lin, C. Y., Imani, V., Saffari, M., Griffiths, M. D., and Pakpour, A. H. (2020). The fear of COVID-19 scale: development and initial validation. *Int. J. Ment. Health Addict.* 27, 1–9. doi: 10.1007/s11469-020-00270-8

Alyami, M., Henning, M., Krägeloh, C. U., and Alyami, H. (2020). Psychometric evaluation of the Arabic version of the fear of COVID-19 scale. *Int. J. Ment. Health Addict.* 16, 1–14. doi: 10.1007/s11469-020-00316-x

Andersen, H. (1973). Abductive and deductive change. *Language* 49, 765–793. doi: 10.2307/412063

Andrich, D. (1988). *Rasch Models for Measurement*. Beverly Hills, LA: Sage Publications.

Arslan, G., Yıldırım, M., Tanhan, A., Buluş, M., and Allen, K. A. (2020). Coronavirus stress, optimism-pessimism, psychological inflexibility, and psychological health: psychometric properties of the coronavirus stress measure. *Int. J. Ment. Health Addict.* 1–17. doi: 10.1007/s11469-020-00337-6 [Epub ahead of print]

Bakker, A. B., and van Wingerden, J. (2021). Rumination about COVID-19 and employee well-being: the role of playful work design. *Can. Psychol.* 62, 73–79. doi: 10.1037/cap0000262

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bitan, D. T., Grossman-Giron, A., Bloch, Y., Mayer, Y., Shiffman, N., and Mendlovic, S. (2020). Fear of COVID-19 scale: psychometric characteristics, reliability and validity in the Israeli population. *Psychiatry Res.* 289:113100. doi: 10.1016/j.psychres.2020.113100

Bond, T. G., and Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences. 2nd Edn*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bretagnolle, J. (2002). "Test of monotonicity for the Rasch model," in *Goodness-of-Fit Tests and Model Validity*. eds. C. Huber-Carol, N. Balakrishnan, M. S. Nikulin and M. Mesbah (Boston, MA: Statistics for Industry and Technology), 365–370.

Burro, R. (2016). To be objective in experimental phenomenology: a psychophysics application. *Springerplus* 5:1720. doi: 10.1186/s40064-016-3418-4

Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J., et al. (2020). The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry Res.* 287:112934. doi: 10.1016/j.psychres.2020.112934

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Modeling* 14, 464–504. doi: 10.1080/10705510701301834

Cheng, C., and Tang, C. S. (2004). The psychology behind the masks: psychological responses to the severe acute respiratory syndrome outbreak in different regions. *Asian J. Soc. Psychol.* 7, 3–7. doi: 10.1111/j.1467-839X.2004.00130.x

Chiu, M. Y. L., Wong, H. T., and Ho, W. W. N. (2020). A comparative study of confirmatory factor analysis and Rasch analysis as item reduction strategies for SAMHSA recovery inventory for Chinese (SAMHSA-RIC). *Eur. J. Psychiatry* 34, 74–81. doi: 10.1016/j.ejpsy.2020.02.002

Christensen, K. B. (2021). Rasch vs. FA. Rasch. Available at: https://www.rasch.org/rmt/rmt263a.htm (Accessed July 30, 2021).

Christensen, K. B., Bjorner, J. B., Kreiner, S., and Petersen, J. H. (2002). Testing unidimensionality in polytomous Rasch models. *Psychometrika* 67, 563–574. doi: 10.1007/BF02295131

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.

Compas, B. E., Connor-Smith, J. K., Saltzman, H., Thomsen, A. H., and Wadsworth, M. (2001). Coping with stress during childhood and adolescence: problems, progress, and potential in theory and research. *Psychol. Bull.* 127, 87–127. doi: 10.1037/0033-2909.127.1.87

Debelak, R., and Koller, I. (2020). Testing the local independence assumption of the Rasch model with Q3-based nonparametric model tests. *Appl. Psychol. Meas.* 44, 103–117. doi: 10.1177/0146621619835501

Deci, E. L., and Ryan, R. M. (1985). The general causality orientations scale: self-determination in personality. *J. Res.* 19, 109–134. doi: 10.1016/0092-6566(85)90023-6

Dorans, N. J., and Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the scholastic aptitude test. *J. Educ. Meas.* 23, 355–368. doi: 10.1111/j.1745-3984.1986.tb00255.x

Dozois, D. J. A.Mental Health Research Canada (2021). Anxiety and depression in Canada during the COVID-19 pandemic: a national survey. *Can. Psychol.* 62, 136–142. doi: 10.1037/cap0000251

Elmer, T., Mepham, K., and Stadtfeld, C. (2020). Students under lockdown: comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. *PLoS One* 15:e0236337. doi: 10.1371/journal.pone.0236337

Estes, K. D., and Thompson, R. R. (2020). Preparing for the aftermath of COVID-19: shifting risk and downstream health consequences. *Psychol. Trauma* 12, S31–S32. doi: 10.1037/tra0000853

Fitzpatrick, K. M., Harris, C., and Drawve, G. (2020). Fear of COVID-19 and the mental health consequences in America. *Psychol. Trauma* 12, S17–S21. doi: 10.1037/tra0000924

Fullana, M. A., Hidalgo-Mazzei, D., Vieta, E., and Radua, J. (2020). Coping behaviors associated with decreased anxiety and depressive symptoms during the COVID-19 pandemic and lockdown. *J. Affect. Disord.* 275, 80–81. doi: 10.1016/j.jad.2020.06.027

Gallagher, M. W., Zvolensky, M. J., Long, L. J., Rogers, A. H., and Garey, L. (2020). The impact of Covid-19 experiences and associated stress on anxiety, depression, and functional impairment in american adults. *Cogn. Ther. Res.* 44, 1043–1051. doi: 10.1007/s10608-020-10143-y

García-Portilla, P., de la Fuente Tomás, L., Bobes-Bascarán, T., Jiménez Treviño, L., Zurrón Madera, P., Suárez Álvarez, M., et al. (2020). Are older adults also at higher psychological risk from COVID-19? *Aging Ment. Health* 25, 1297–1304. doi: 10.1080/13607863.2020.1805723

Hagquist, C., and Andrich, D. (2017). Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health Qual. Life Outcomes* 15:181. doi: 10.1186/s12955-017-0755-0

Haktanir, A., Seki, T., and Dilmaç, B. (2020). Adaptation and evaluation of Turkish version of the fear of COVID-19 scale. *Death Stud.* 1–9. doi: 10.1080/07481187.2020.1773026 [Epub ahead of print]

Hastings, P. D., Miller, J. G., and Troxel, N. R. (2015). "Making good: the socialization of children's prosocial development," in *Handbook of Socialization: Theory and Research*. eds. J. E. Grusec and P. D. Hastings (New York, NY: The Guilford Press), 637–660.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179–185. doi: 10.1007/BF02289447

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Husky, M. M., Kovess-Masfety, V., and Swendsen, J. D. (2020). Stress and anxiety among university students in France during Covid-19 mandatory confinement. *Compr. Psychiatry* 102:152191. doi: 10.1016/j.comppsych.2020.152191

Ishibashi, R., Nouchi, R., Honda, A., Abe, T., and Sugiura, M. (2019). A concise psychometric tool to measure personal characteristics for surviving natural disasters: development of a 16-item power to live questionnaire. *Geosciences* 9:366. doi: 10.3390/geosciences9090366

Kang, H. A., Su, Y. H., and Chang, H. H. (2018). A note on monotonicity of item response functions for ordered polytomous item response theory models. *Br. J. Math. Stat. Psychol.* 71, 523–535. doi: 10.1111/bmsp.12131

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling. 4th Edn*. New York: Guilford Press.

Kreiner, S., and Christensen, K. B. (2013). "Person parameter estimation and measurement in Rasch models," in *Rasch Models in Health*. eds. K. B. Christensen, S. Kreiner and M. Mesbah (London, UK, Hoboken, NJ: ISTE Ltd., Wiley & Sons), 63–78.

Lazarus, R. S., and Folkman, S. (1984). *Stress, Appraisal, and Coping*. New York, NY: Springer.

Lee, S. A. (2020). Coronavirus anxiety scale: A brief mental health screener for COVID-19 related anxiety. *Death Stud.* 44, 393–401. doi: 10.1080/07481187.2020.1748481

Lenth, R. V. (2021). emmeans: estimated marginal means, aka least-squares means. R Package Version 1.6.2-1. Available at: https://CRAN.R-project.org/package=emmeans (Accessed July 8, 2021).

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *J. Appl. Meas.* 3, 85–106.

Lüdecke, D., Ben-Shachar, M., Patil, I., and Makowski, D. (2020). Extracting, computing and exploring the parameters of statistical models using R. *J. Open Source Soft.* 5:2445. doi: 10.21105/joss.02445

Mair, P., Hatzinger, R., and Maier, M. J. (2021). eRm: extended rasch modeling. R Package Version 1.0-2. Available at: https://CRAN.R-project.org/package=eRm (Accessed July 8, 2021).

Marais, I. (2013). "Local dependence," in *Rasch Models in Health*. eds. K. B. Christensen, S. Kreiner and M. Mesbah (London, UK, Hoboken, NJ: ISTE Ltd., Wiley & Sons), 111–130.

Marsh, H. W., Hau, K.-T., and Grayson, D. (2005). "Goodness of fit evaluation in structural equation modeling," in *Contemporary Psychometrics*. eds. A. Maydeu-Olivares and J. McArdle (Mahwah, NJ: Erlbaum), 275–340.

Masters, G. N., and Wright, B. D. (1997). "The partial credit model," in *Handbook of Modern Item Response Theory*. eds. LindenW. J. Van der and R. K. Hambleton (New York, NY: Springer), 101–121.

Nicomedes, C., and Avila, R. (2020). An analysis on the panic during COVID-19 pandemic through an online form. *J. Affect. Disord.* 276, 14–22. doi: 10.1016/j.jad.2020.06.046

Odriozola-González, P., Planchuelo-Gómez, Á., Irurtia, M. J., and de Luis-García, R. (2020). Psychological effects of the COVID-19 outbreak and lockdown among students and workers of a Spanish university. *Psychiatry Res.* 290:113108. doi: 10.1016/j.psychres.2020.113108

Panayiotou, G., Panteli, M., and Leonidou, C. (2021). Coping with the invisible enemy: the role of emotion regulation and awareness in quality of life during the COVID-19 pandemic. *J. Contextual Behav. Sci.* 19, 17–27. doi: 10.1016/j.jcbs.2020.11.002

Panella, L., La Porta, F., Caselli, S., Marchisio, S., and Tennant, A. (2012). Predicting the need for institutional care shortly after admission to rehabilitation: Rasch analysis and predictive validity of the BRASS index. *Eur. J. Phys. Rehabil. Med.* 48, 443–454.

Park, C. L., Finkelstein-Fox, L., Russell, B. S., Fendrich, M., Hutchison, M., and Becker, J. (2021). Americans' distress early in the COVID-19 pandemic: protective resources and coping strategies. *Psychol. Trauma* 13, 422–431. doi: 10.1037/tra0000931

Pekrun, R., and Bühner, M. (2014). "Self-report measures of academic emotions," in *Educational Psychology Handbook Series. International Handbook of Emotions in Education*. eds. R. Pekrun and L. Linnenbrink-Garcia (London, UK: Routledge/Taylor & Francis Group), 561–579.

Porcelli, P. (2020). Fear, anxiety and health-related consequences after the COVID-19 epidemic. *Clin.* 17, 103–111. doi: 10.36131/CN20200215

R Core Team (2021). R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available at: https://www.R-project.org/ (Accessed July 8, 2021).

Raccanello, D., Barnaba, V., Rocca, E., Vicentini, G., Hall, R., and Burro, R. (2021a). Adults' expectations on children's earthquake-related emotions and coping strategies. *Psychol. Health Med.* 26, 571–583. doi: 10.1080/13548506.2020.1800057

Raccanello, D., Brondino, M., Crescentini, A., Castelli, L., and Calvo, S. (2021b). A brief measure for school-related achievement emotions: the achievement emotions adjective list (AEAL) for secondary students. *Eur. J. Dev. Psychol.* 1–19. doi: 10.1080/17405629.2021.1898940 [Epub ahead of print]

Raccanello, D., Rocca, E., and Brondino, M. (2019). "Disaster-related coping strategies: a meta-analysis on children. in *19th European Conference on Developmental Psychology*. Abstract Book, ed. National and Kapodistrian University of Athens (Athens, Greece: Global Events), 694.

Raccanello, D., Vicentini, G., Florit, E., and Burro, R. (2020a). Factors promoting learning with a web application on earthquake-related emotional preparedness in primary school. *Front. Psychol.* 11:621. doi: 10.3389/fpsyg.2020.00621

Raccanello, D., Vicentini, G., Rocca, E., Barnaba, V., Hall, R., and Burro, R. (2020b). Development and early implementation of a public communication campaign to help adults to support children and adolescents to cope with coronavirus-related emotions: a community case study. *Front. Psychol.* 11:2184. doi: 10.3389/fpsyg.2020.02184

Raccanello, D., Vicentini, G., Trifiletti, E., and Burro, R. (2021c). A Rasch analysis of the school-related well-being (SRW) scale: measuring well-being in the transition from primary to secondary school. *Int. J. Environ. Res. Public Health* 18:23. doi: 10.3390/ijerph18010023

Raiche, G., and Magis, D. (2020). nFactors: parallel analysis and other non graphical solutions to the cattell scree test. R Package Version 2.4.1. Available at: https://CRAN.R-project.org/package=nFactors

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago, IL: The University of Chicago Press.

Revelle, W. (2021). psych: procedures for personality and psychological research. R Package Version 2.1.6. Available at: https://CRAN.R-project.org/package=psych (Accessed July 8, 2021).

Revelle, W., and Rocklin, T. (1979). Very simple structure: an alternative procedure for estimating the optimal number of interpretable factors. *Multivar. Behav. Res.* 14, 403–414. doi: 10.1207/s15327906mbr1404_2

Reznik, A., Gritsenko, V., Konstantinov, V., Khamenka, N., and Isralowitz, R. (2020). COVID-19 fear in eastern Europe: validation of the fear of COVID-19 scale. *Int. J. Ment. Health Addict.* 1–6. doi: 10.1007/s11469-020-00283-3 [Epub ahead of print]

Rodríguez-Rey, R., Garrido-Hernansaiz, H., and Collado, S. (2020). Psychological impact of COVID-19 in Spain: early data report. *Psychol. Trauma* 12, 550–552. doi: 10.1037/tra0000943

Rönkkö, M., and Cho, E. (2020). An updated guideline for assessing discriminant validity. *Organ. Res. Methods* 1:42. doi: 10.1177/1094428120968614

Rosenfield, S. (2000). "Gender and dimensions of the self: implications for internalizing and externalizing behavior," in *Gender and its Effects on Psychopathology*. ed. E. Frank (Washington, DC: American Psychiatric Publishing, Inc.), 23–36.

Rosseel, Y. (2012). Lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

Ruscio, J., and Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychol. Assess.* 24, 282–292. doi: 10.1037/a0025697

Sakib, N., Bhuiyan, A., Hossain, S., Al Mamun, F., Hosen, I., Abdullah, A. H., et al. (2020). Psychometric validation of the Bangla fear of COVID-19 scale: confirmatory factor analysis and Rasch analysis. *Int. J. Ment. Health Addict.* 1–12. doi: 10.1007/s11469-020-00289-x [Epub ahead of print]

Satici, B., Saricali, M., Satici, S. A., and Griffiths, M. D. (2020). Intolerance of uncertainty and mental wellbeing: serial mediation by rumination and fear of COVID-19. *Int. J. Ment. Health Addict.* 1–12. doi: 10.1007/s11469-020-00305-0 [Epub ahead of print]

Schaefer, J. A., and Moos, R. H. (1992). "Life crises and personal growth," in *Personal Coping: Theory, Research, and Application*. ed. B. N. Carpenter (Westport, CT: Praeger), 149–170.

Shamblaw, A. L., Rumas, R. L., and Best, M. W. (2021). Coping during the COVID-19 pandemic: relations with mental health and quality of life. *Can. Psychol.* 62, 92–100. doi: 10.1037/cap0000263

Simons, D. J., Shoda, Y., and Lindsay, D. S. (2017). Constraints on generality (COG): a proposed addition to all empirical papers. *Perspect. Psychol. Sci.* 12, 1123–1128. doi: 10.1177/1745691617708630

Skinner, E. A., Edge, K., Altman, J., and Sherwood, H. (2003). Searching for the structure of coping: a review and critique of category systems for classifying ways of coping. *Psychol. Bull.* 129, 216–269. doi: 10.1037/0033-2909.129.2.216

Skinner, E. A., and Zimmer-Gembeck, M. J. (2007). The development of coping. *Annu. Rev. Psychol.* 58, 119–144. doi: 10.1146/annurev.psych.58.110405.085705

Smith, E. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J. Appl. Meas.* 3, 205–231.

Son, C., Hegde, S., Smith, A., Wang, X., and Sasangohar, F. (2020). Effects of COVID-19 on college students' mental health in the United States: interview survey study. *J. Med. Internet Res.* 22:e21279. doi: 10.2196/21279

Soraci, P., Ferrari, A., Abbiati, F. A., Del Fante, E., De Pace, R., Urso, A., et al. (2020). Validation and psychometric evaluation of the Italian version of the fear of COVID-19 scale. *Int. J. Ment. Health Addict.* 1–10. doi: 10.1007/s11469-020-00277-1 [Epub ahead of print].

Sugiura, M., Sato, S., Nouchi, R., Honda, A., Abe, T., Muramoto, T., et al. (2015). Eight personal characteristics associated with the power to live with disasters as indicated by survivors of the 2011 great East Japan earthquake disaster. *PLoS One* 10:e0130349. doi: 10.1371/journal.pone.0130349

Tang, W., Hu, T., Yang, L., and Xu, J. (2020). The role of alexithymia in the mental health problems of home-quarantined university students during the COVID-19 pandemic in China. *Pers. Individ. Differ.* 165:110131. doi: 10.1016/j.paid.2020.110131

Tennant, A., and Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care Res.* 57, 1358–1362. doi: 10.1002/art.23108

Tennant, A., Penta, M., Tesio, L., Grimby, G., Thonnard, J. L., Slade, A., et al. (2004). Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. *Med. Care* 42, 137–148. doi: 10.1097/01.mlr.0000103529.63132.77

Vicentini, G., Brondino, M., Burro, R., and Raccanello, D. (2020). HEMOT®, helmet for EMOTions: a web application for children on earthquake-related emotional prevention. *Adv. Intell. Systems Comp.* 1241, 10–19. doi: 10.1007/978-3-030-52538-5_2

Vidotto, G., Moroni, L., Burro, R., Filipponi, L., Balestroni, G., Bettinardi, O., et al. (2010). A revised short version of the depression questionnaire. *Eur. J. Cardiovasc. Prev. Rehabil.* 17, 187–197. doi: 10.1097/HJR.0b013e328333edc8

Vinkers, C. H., van Amelsvoort, T., Bisson, J. I., Branchi, I., Cryan, J. F., Domschke, K., et al. (2020). Stress resilience during the coronavirus pandemic. *Eur. Neuropsychopharmacol.* 35, 12–16. doi: 10.1016/j.euroneuro.2020.05.003

Wakashima, K., Asai, K., Kobayashi, D., Koiwa, K., Kamoshida, S., and Sakuraba, M. (2020). The Japanese version of the fear of COVID-19 scale: reliability, validity, and relation to coping behavior. *PLoS One* 15:e0241958. doi: 10.1371/journal.pone.0241958

Waselewski, E. A., Waselewski, M. E., and Chang, T. (2020). Needs and coping behaviors of youth in the U.S. during COVID-19. *J. Adolesc. Health* 67, 649–652. doi: 10.1016/j.jadohealth.2020.07.043

Winter, T., Riordan, B. C., Pakpour, A. H., Griffiths, M. D., Mason, A., Poulgrain, J. W., et al. (2020). Evaluation of the English version of the fear of COVID-19 scale and its relationship with behavior change and political beliefs. *Int. J. Ment. Health Addict.* 1–11. doi: 10.1007/s11469-020-00342-9 [Epub ahead of print]

Wright, B. D. (2001). Separation, reliability and skewed distributions: statistically different levels of performance. *Rasch Meas. Trans.* 14:786.

Wright, B. D., and Linacre, J. M. (1994). Reasonable mean-square-fit values. *Rasch Meas. Trans.* 8:370.

Wright, B. D., and Masters, G. N. (1982). *Rating Scale Analysis*. Chicago, IL: MESA Press.

Zimmer-Gembeck, M. J., and Skinner, E. A. (2011). The development of coping across childhood and adolescence: an integrative review and critique research. *Int. J. Behav. Dev.* 35, 1–17. doi: 10.1177/0165025410384923

Conflict of Interest: Author Rob Hall is employed by Environmetrics Pty Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 September 2021
doi: 10.3389/fpsyg.2021.724187

Check for updates

Validation of the Chinese Version of the Self-Objectification Beliefs and Behaviors Scale

*Min Lang and Yiduo Ye**

School of Psychology, Fujian Normal University, Fuzhou, China

Given the limitations of the existing tools used for measuring self-objectification in China, this study aims to validate the Chinese version of the self-objectification beliefs and behaviors scale (C-SOBBS). In this study, we first translated and culturally adopted SOBBS to the Chinese context. We conducted two wave surveys. In the first-wave survey, we recruited 331 female college students whose age ranged from 18 to 35 (M_{age} = 20.28, SD = 2.99) to complete an online survey that included demographic questions, C-SOBBS, and four other scales to assess the validity of C-SOBBS. In the second-wave survey, 76 participants who took part in the first-wave survey completed the C-SOBBS at a two-week interval for the assessment of test-retest stability. A confirmatory factor analysis was performed to validate the factor structure of the C-SOBBS. The relationship between the C-SOBBS, its factors, and four other measures demonstrated that the C-SOBBS has a convergent and discriminant validity. Furthermore, the results of hierarchical multiple regression demonstrated the C-SOBBS's incremental validity related to the Female Questionnaire of Trait Self-Objectification and Objectified Body Consciousness-Surveillance subscale. Additionally, the internal consistency and test-retest reliability of the C-SOBBS were also verified. The results of this study demonstrate the utility of the C-SOBBS in assessing the self-objectification beliefs and behaviors of young Chinese women within the context of Chinese culture.

Keywords: self-objectification, young Chinese women, measurement, validity, reliability

OPEN ACCESS

Edited by:
Paola Gremigni,
University of Bologna, Italy

Reviewed by:
Jian Li,
Beijing Normal University, China
Antonio Granero-Gallegos,
University of Almeria, Spain

****Correspondence:***
Yiduo Ye
yeyiduo@163.com

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *12 June 2021*
Accepted: *27 August 2021*
Published: *24 September 2021*

Citation:
Lang M and Ye Y (2021) Validation of the Chinese Version of the Self-Objectification Beliefs and Behaviors Scale.
Front. Psychol. 12:724187.
doi: 10.3389/fpsyg.2021.724187

INTRODUCTION

Bartky (1990) proposed the concept of sexual objectification and defined it as "the separation of one's body, body parts, and sexual functions from one's identity, thereby reducing a person to the status of an object." Fredrickson and Roberts (1997) argued that western culture is full of sexual objectification, such as close-ups of women's sexual body parts in various forms of visual media and men treating women as sexual objects even in interpersonal situations. Women exposed to this kind of influence for a prolonged period will gradually accept and internalize these objectifying attitudes and view their own body from a third person's perspective, which leads to self-objectification. According to objectification theory, women practicing self-objectification habitually monitor their own physical appearance (Fredrickson and Roberts, 1997), which leads to body image issues, such as body shame (Fredrickson and Roberts, 1997; Adams et al., 2017; Schaefer et al., 2018; Baildon et al., 2021), body dissatisfaction

(Grippo and Hill, 2008; Schaefer et al., 2018), physical anxiety (Fredrickson and Roberts, 1997; Jongenelis et al., 2014; Adams et al., 2017), disordered eating (Cohen et al., 2018; Schaefer and Thompson, 2018; Kilpela et al., 2019), depression (Jones and Griffiths, 2015; Register et al., 2015; Vencill et al., 2015), and sexual dysfunction (Fredrickson and Roberts, 1997; Tiggemann, 2011). Physical shame as a common consequence of self-objectification has been extensively studied. For example, Choma et al. (2009) found that self-objectification can predict body shame among Canadian undergraduate women. Similarly, the predictive effect of self-objectification on body shame has also been observed among young Chinese women (Sun and Zheng, 2016; Teng et al., 2019; Wang et al., 2020).

In the two decades after Fredrickson and Roberts (1997) proposed the construct of self-objectification, five self-report scales were devised to assess it, namely, the Self-Objectification Questionnaire (SOQ; Fredrickson et al., 1998), the Body Surveillance Subscale of the Objectified Body Consciousness Scale (OBC-Surveillance; McKinley and Hyde, 1996), the Self-Objectification Scale (SOS; Talmon and Ginzburg, 2016), the Self-Objectification Beliefs and Behaviors Scale (SOBBS; Lindner and Tantleff-Dunn, 2017), and the Female Questionnaire of Trait Self-Objectification (FQSO; Wu and Lang, 2019).

In the SOQ, self-objectification is defined as valuing observable physical appearance (e.g., "How do I look?") over non-observable physical competence (e.g., "What am I capable of?"). The SOQ items include 10 body attributes, including five appearance-based attributes (e.g., weight) and five competence-based attributes (e.g., health; Fredrickson et al., 1998). Participants were asked to rank the 10 body attributes from 0 = *least impact on my physical self-concept* to 9 = *great impact on my physical self-concept*. To obtain a final score, the sum of the ranks assigned to the five competence-based attribute scores is subtracted from the sum of the ranks given to appearance-based scores (Fredrickson et al., 1998). A score greater than 0 indicates a greater emphasis on physical appearance, while a score less than 0 indicates a greater emphasis on physical competence (Fredrickson et al., 1998). Although the SOQ is widely used, there are still some shortcomings, mainly in the following aspects. First, the item ranking method forces participants to put physical competence in direct opposition to physical appearance (Hill and Fischer, 2008; Calogero, 2011); however, as a stable trait, at least two other situations may also exist: valuing both physical competence and physical appearance, and valuing neither physical competence nor physical appearance (Wu and Lang, 2019). Second, the item rating method makes it difficult for participants to complete the SOQ correctly; therefore, many questionnaires are unsuitable for data analysis (Calogero, 2011; Lindner and Tantleff-Dunn, 2017; Wu and Lang, 2019). Third, the SOQ's internal consistency coefficient cannot be calculated because of its rank-order format and scoring system (Hill and Fischer, 2008).

The OBC-Surveillance is a subscale of the Objectified Body Consciousness Scale (OBCS) developed by McKinley and Hyde in 1996. The OBC-Surveillance is mainly used to measure the degree and extent to which a woman perceives herself as an external observer would (McKinley and Hyde, 1996). It includes eight items (e.g., "I think more about how my body feels than how my body looks"). According to objectification theory, self-objectification may cause self-surveillance, which then leads to negative attitudes toward oneself, such as body dissatisfaction and body shame (Fredrickson and Roberts, 1997). Although self-objectification and self-surveillance are similar in concept, different scholars have different views on whether they are interchangeable or equivalent (Calogero, 2011; Lindner and Tantleff-Dunn, 2017). Tiggemann and Kuring (2004) used the SOQ score to predict the OBC-Surveillance score, surmising that self-surveillance is an expected outcome of self-objectification. Some argue that body surveillance is equivalent to self-objectification, so they used OBC-Surveillance alone to measure the latter (Augustus-Horvath and Tylka, 2009; Lindner et al., 2012). In short, it is not clear whether self-surveillance can fully reflect self-objectification (Lindner and Tantleff-Dunn, 2017). The definition of self-objectification includes not only the third-party perspective of the observer but also the excessive value placed on physical appearance over physical competence, as well as treating the body as if it is the only thing that represents the self (Lindner and Tantleff-Dunn, 2017). Thus, OBC-Surveillance is inadequate and inappropriate for measuring self-objectification.

The SOS was developed to assess the nullifying experience of self-objectification by Talmon and Ginzburg (2016). In the SOS, self-objectification is defined as "a state in which individuals perceive themselves as objects and instruments to satisfy the needs and desires of others" (Talmon and Ginzburg, 2016). It includes two factors: invisibility and lack of autonomy, and 17 items (e.g., "Many times people ignore my feelings"). The total score was obtained by averaging all items. Higher scores indicated higher self-objectification (Talmon and Ginzburg, 2016). Unlike the SOQ and the OBC-Surveillance, which assess self-objectification as representing self-perception based on sexual and bodily appearance, the SOS assesses self-objectification as reflecting dehumanization (Talmon and Ginzburg, 2016). Therefore, the SOS cannot be used to measure self-objectification, which focuses on physical appearance.

The SOBBS is a 14-item measure with a 5-point response format. It comprises two factors: the observer's perspective ("thinking the body as an observer would") and the body as self ("treating the body as if it is capable of representing the self as a person"; Lindner and Tantleff-Dunn, 2017, p. 256). The total score was obtained by averaging the item scores. Higher scores indicated higher self-objectification (Lindner and Tantleff-Dunn, 2017). The SOBBS has several advantages. The first and most important one is that the conceptual definition of self-objectification is more complete (Lindner and Tantleff-Dunn, 2017). Lindner and Tantleff-Dunn (2017) integrated conceptual and operational definitions of self-objectification and sexual objectification into a single measure. Therefore, the SOBBS can measure both the internalized observer's perspective (as was done in OBC-Surveillance), the value of physical appearance over physical competence (as was done in the SOQ), and treating the body as if it is the sole representation of the self (as is highlighted by the definition of sexual objectification). Second, each item of the SOBBS is a statement;

this form is better understood than the body attributes used by the SOQ (Wu and Lang, 2019). Third, Lindner and Tantleff-Dunn (2017) found that the SOBBS can better predict physical shame and appearance anxiety than the SOQ and OBC-Surveillance.

Based on the SOQ, Wu and Lang (2019) designed the FQSO specifically to measure self-objectification in Chinese women. It includes 17 items, 10 about body-appearance attributes (e.g., facial features and facial shape) and 7 body-competence attributes (e.g., body flexibility). Unlike the SOQ, the FQSO uses a 7-point Likert scale for scoring. To some extent, it overcomes some shortcomings of the SOQ, including its inability to calculate internal consistency reliability and a high rate of incomplete questionnaires. However, there are still some shortcomings. For example, some participants thought the test was unnecessary because they believed that all body attributes were important (Wu and Lang, 2019). Some believed that certain items (e.g., health) were important to all people, and some participants found it confusing to use words as items. Furthermore, when evaluating self-objectification, the FQSO still adopts the definition of "valuing physical appearance more than physical competence," which is similar to the SOQ. This definition may not fully summarize the complexity of self-objectification; therefore, underlying structural problems still exist in the questionnaire.

Overall, all five measurements were used to assess self-objectification and reported satisfactory reliability and validity (McKinley and Hyde, 1996; Fredrickson et al., 1998; Talmon and Ginzburg, 2016; Lindner and Tantleff-Dunn, 2017; Wu and Lang, 2019). Although the SOQ and the OBC-Surveillance are the most commonly used scales, they are also the most criticized ones (Calogero, 2011; Wu and Lang, 2019). The SOS is not suitable for measuring self-objectification concerning appearance (Talmon and Ginzburg, 2016). The FQSO, although developed specifically for measuring self-objectification in Chinese women, cannot fully measure all the connotations of self-objectification, focusing on appearance (Wu and Lang, 2019). The SOBBS is the most suitable tool for fully assessing self-objectification, which focuses on appearance. However, there is no Chinese version of the SOBBS. Hence, this study intends to adopt and validate the SOBBS to suit the Chinese context.

The Present Study

This study aims to establish a Chinese version of the SOBBS (C-SOBBS) and to verify the psychometric properties of the translated scale. First, we translate and culturally adopt SOBBS for the Chinese context. Second, we conduct a confirmatory factor analysis (CFA) to evaluate the factor structure of the C-SOBBS. We then evaluate the scale's convergent and discriminant validity through its relations with the FQSO, body surveillance, body shame, and sexual objectification. Next, we evaluate the incremental validity of the C-SOBBS relative to the FQSO and OBC-Surveillance. Finally, we establish the test-retest reliability of the C-SOBBS using a smaller sample of 76 adult Chinese women. The specific hypotheses for this study are as follows: (1) The C-SOBBS would demonstrate the best fit through CFA for a sample of young Chinese women. (2) The "observer's perspective" factor would be moderately correlated with the "body as self" factor, and both would be highly correlated with the C-SOBBS total score. (3) The C-SOBBS and its factors are positively correlated with the FQSO and OBC-Surveillance. (4) The C-SOBBS and its factors are positively correlated with body shame and sexual objectification. (5) The C-SOBBS predicts body shame above the FQSO and OBC-Surveillance. (6) The first-wave survey (T1) of the C-SOBBS and its factors would be highly correlated with the second-wave survey (T2) of the C-SOBBS and its factors.

MATERIALS AND METHODS

Translation and Cultural Adaptation of the SOBBS to the Chinese Context

According to the guidelines for cross-cultural adaptation of instruments (Beaton et al., 2000; Swami and Barron, 2019), we first obtained the original English version of the SOBBS and then contacted three professional translators to translate it. They first translated SOBBS independently; after the translation, they held a discussion to get a Chinese version of the SOBBS that all of them agreed on. We then asked two other translators to translate the text back into English. After that, we invited two translators and three experts in self-objectification to discuss the semantic, habitual, cultural, and conceptual equivalence between the original and the Chinese version of SOBBS, named C-SOBBS 1. Next, we invited eight people who studied body image and scale development to provide feedback about the equivalence between the original and C-SOBBS 1 on a 7-point scale (1 = not at all and 7 = completely equivalent). If the mean of the item was less than 6, we modified it until the mean reached 6 and thus arrived at C-SOBBS 2. Thereafter, we interviewed five university students who were not psychology students and asked them to give feedback on the sentence intelligibility and semantics of 14 items. Finally, we obtained C-SOBBS 3 based on feedback and discussion with two experts.

Participants and Procedure

Human participation in this study was reviewed and approved by the Fujian Normal University Ethics Committee. For data collection, we adopted convenience sampling and conducted the survey twice. In the first round, we developed an online survey that included demographic questions, C-SOBBS 3, and four other scales through the WJX.cn platform. Then, we contacted some colleagues who taught public courses at universities to help us recruit 339 female college students to complete the online survey between classes. All participants signed an informed consent form before completing the questionnaire. They were also told to complete the survey within 10 min. After completing the questionnaire, we gave small gifts as compensation and invited participants who were willing to participate in the retest to provide their contact information. A total of 100 participants left their contact information, and 76 of them participated in the second survey 2 weeks later. All data in the second survey were used to analyze the test-retest reliability. In the first round, only 331 correctly completed surveys were used for data analysis because eight surveys were deleted due to the completion time being too short (<5 min) and/or incorrect answers to items with specified options. Participants' age ranged from 18 to 35 years ($M_{age}=20.28$, $SD=2.99$). The participants were mainly recruited from a university in Zigong City, Sichuan Province and a university in Fuyang City, Anhui Province. The CFA model evaluated in this study has 76 degrees of freedom. A sample size of 331 was sufficient for the analyses based on the sample size guidelines published by MacCallum et al. (1996). The test-retest sample size ($n=76$) in this study was larger than that ($n=55$) in the original SOBBS study. This demonstrates that a sample size of 76 was sufficient.

MEASURES

The Chinese Version of Self-Objectification Beliefs and Behaviors Scale

The C-SOBBS, translated and culturally adapted, was used to measure self-objectification beliefs and behaviors among young Chinese women. The scale consists of 14 items [e.g., "I try to imagine what my body looks like to others (i.e., like I am looking at myself from the outside)"]. Participants rated each of the 14 items on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). The overall score was obtained by averaging all the item scores. Higher scores indicate stronger self-objectification beliefs and more self-objectification behaviors. The internal consistency of the SOBBS was 0.91 (Lindner and Tantleff-Dunn, 2017). The satisfactory construct validity of the SOBBS was originally reported by Lindner and Tantleff-Dunn (2017).

The Female Questionnaire of Trait Self-Objectification

The FQSO (Wu and Lang, 2019) was used to measure the degree of trait self-objectification in Chinese women. It includes two factors: physical appearance and physical competence. It has 17 body attributes (e.g., hair, eyes, height, body flexibility, and strength). Participants were asked to rate each item on a 7-point scale, ranging from 1 (least important) to 7 (most important). The overall score was derived by subtracting the competency-based score from the appearance-based score. Higher scores indicated higher trait self-objectification. For Chinese female undergraduate students, Wu and Lang (2019) reported that the internal consistency reliability of the physical appearance factor was 0.89, and the physical competency factor was 0.82. In the current study, Cronbach's alpha was 0.87 for the factor of physical appearance and 0.89 for the factor of physical competencies. Wu and Lang (2019) reported satisfactory construct validity for a sample of young Chinese women.

The Body Surveillance Subscale (OBC-Surveillance)

The OBC-Surveillance assesses the degree to which women assume an observer's perspective of their bodies. It comprises eight items (e.g., "I often worry about whether the clothes I am wearing make me look good"). Respondents rated each of the eight items on a 7-point response scale ranging from 1 (strongly disagree) to 7 (strongly agree). The total score was obtained by summing the item scores. Higher scores indicated a stronger tendency to take an outsider's perspective while viewing their own body. For Chinese undergraduate students, the internal consistency reliability of the OBC-Surveillance was 0.88 (Chen and Jiang, 2007). For American young women and middle-aged women, the internal consistency reliability of the OBC-Surveillance was 0.89 (McKinley and Hyde, 1996). In this study, the internal consistency of the OBC-Surveillance was 0.81. McKinley and Hyde (1996) originally reported the construct validity of the OBC, and the satisfactory construct validity of the Chinese version of the OBC-Surveillance scale was also demonstrated by Chen and Jiang (2007).

The Body Shame Subscale

The BSS is a subscale of the OBCS (McKinley and Hyde, 1996). It assesses the degree to which women feel that they are bad people when they view themselves as not meeting cultural appearance standards, particularly thinness. It comprises eight items (e.g., "When I can't control my weight, I feel like something must be wrong with me"). Respondents rated each of the eight items on a 7-point response scale ranging from 1 (strongly disagree) to 7 (strongly agree), as in McKinley and Hyde (1996). The total score was obtained by summing the item scores. Higher scores indicated higher body shame. For Chinese undergraduate students, the internal consistency reliability of the OBC-BSS was 0.86 (Chen and Jiang, 2007). Schaefer et al. (2018) reported similar reliability estimates (white college women: $\alpha=0.80$; black college women: $\alpha=0.75$). In the current study, the internal consistency of the BSS was 0.82, and McKinley and Hyde (1996) originally reported the construct validity of the OBC, and satisfactory construct validity of the Chinese version of the BSS was also demonstrated by Chen and Jiang (2007).

The Interpersonal Sexual Objectification Scale

The ISOS is used to measure the frequency of sexual objectification experienced within the past year (Kozee et al., 2007). It comprises 15 items (e.g., "How often have you noticed someone staring at your breasts when you were talking to them?") with a 5-point response format (ranging from 1 = never to 5 = always). The total score was obtained by summing the item scores. Higher scores indicated greater experience of sexual objectification. For Chinese undergraduate students, the internal consistency reliability of the ISOS was 0.87 (Sun and Zheng, 2016). In the current study, Cronbach's alpha is 0.91, and satisfactory construct validity was originally demonstrated by Kozee et al. (2007). Sun and Zheng (2016) also reported satisfactory construct validity in the Chinese version.

Data Analyses

We examined the factor structure of the C-SOBBS using CFA performed in AMOS 22.0. To evaluate the model fit, widely accepted fit indices were used as: the goodness-of-fit index (GFI; near or above 0.90), comparative fit index (CFI; near or above 0.95), the ratio of the chi-square to the degree of freedom (χ^2/df), and the root mean square error of approximation (RMSEA; at or below 0.06) with a 90% confidence interval (CI; Hu and Bentler, 1999). The validity was performed in three ways: (a) Pearson's correlations were used among the two factors of C-SOBBS, overall C-SOBBS, and four other scales, (b) a set of estimations of composite reliability (CR) and average variance extracted (AVE) of each factor were examined, and (c) a hierarchical multiple regression was conducted to analyze the incremental validity of the C-SOBBS relative to the FQSO and OBC-Surveillance. We assessed internal consistency reliability with Cronbach's alpha and test-retest reliability with the intraclass correlation coefficient (ICC) using the commonly reported cutoff values of 0.70 and 0.80 (Nunnally, 1978; Fleiss, 1999).

RESULTS

Descriptive and Correlation Statistics

As presented in **Table 1**, the mean for 14 items of C-SOBBS ranged from 2.05 to 3.23, and the standard deviation (SD) for 14 items of C-SOBBS ranged from 0.81 to 1.04. For both dimensions, all items had values of kurtosis (ranges from −0.79 to 0.75) and skewness (ranges from −0.40 to 0.88), indicative of slight deviations from the normal distribution. According to the literature review by Kline (1998), there are no sensitivity problems or significant non-normality. There were significant correlations among 14 items of C-SOBBS which ranged from 0.23 to 0.58 ($p < 0.01$).

Validation of the C-SOBBS's Factor Structure

A CFA with maximum likelihood estimation was conducted to confirm the two-factor structure of the 14-item C-SOBBS. The model provided an excellent fit for the sample, GFI = 0.93, CFI = 0.95, RMSEA = 0.06 [90% CI (0.048, 0.073)], and χ^2/df was statistically significant for the model, χ^2/df = 2.19, $p < 0.001$. To prove that the two-factor model would be better than the single-factor model, we also conducted a CFA to confirm the one-factor structure of the 14-item C-SOBBS, GFI = 0.84, CFI = 0.86, RMSEA = 0.10. These results indicate that the two-factor C-SOBBS demonstrated the best fit for the young Chinese women sample, thereby supporting Hypothesis 1. The standardized regression weights of each item on the corresponding latent factor were all significant and valued between 0.56 and 0.79 (**Table 2**).

Validity

As presented in **Table 3**, both the "observer's perspective" factor and the "body as self" factor had significant positive correlations with overall C-SOBBS ($r = 0.92$ and $r = 0.89$, $p < 0.01$), and the correlation between the factors was 0.64 ($p < 0.01$). There was a significant moderate correlation between the two factors and a significantly high correlation between each factor and the overall C-SOBBS. Further, the correlation between the two factors was lower than the correlation between the two factors and the overall C-SOBBS, which supports Hypothesis 2. The composite reliability (CR) of two factors of C-SOBBS was 0.86 and 0.84, and the AVE was 0.47 and 0.44, respectively. The data show that acceptable reliability indices were achieved (Fornell and Larcker, 1981), and the minimum AVE (0.44) was greater than the square of the correlation coefficient of the two factors (0.41). The results show that the C-SOBBS has high convergent and discriminant validity. Furthermore, the two factors and the overall C-SOBBS had significant positive correlations with the FQSO and the OBC-Surveillance, supporting Hypothesis 3. This supports the convergent validity of C-SOBBS. In addition, the "observer's perspective" factor, the "body as self" factor, and the overall C-SOBBS had significant positive correlations with BSS and ISOS, supporting Hypothesis 4, indicating that the C-SOBBS has good criterion validity. In this study, we also conducted hierarchical multiple regression to analyze whether the C-SOBBS predicted body shame beyond the two measures used to assess self-objectification in China—the FQSO and OBC-Surveillance. As presented in **Table 4**, the FQSO predicted the body shame in Model 1. In model 2, both FQSO and OBC-Surveillance predicted body shame. In model 3, the FQSO and C-SOBBS predicted body shame, whereas the OBC-Surveillance did not. Collectively, the addition of C-SOBBS to the regression model results in a significant improvement in the prediction of body shame. Thus, Hypothesis 5 is validated. Taken together, these results provide strong evidence for the validity of C-SOBBS.

Reliability

The final 14-item C-SOBBS shows excellent internal consistency. Cronbach's alpha was 0.86 for the "observer's perspective" factor and 0.84 for the "body as self" factor, and the overall internal consistency of C-SOBBS was 0.90. To establish the test-retest reliability of the C-SOBBS over a two-week interval, 76 participants completed the survey. The results of the ICC (3.1) indicate that the test-retest reliability of the overall C-SOBBS (0.86), the observer's perspective factor (0.74), and the factor of the "body as self" (0.72) were excellent. Thus, Hypothesis 6 is validated.

General Discussion

The primary goal of our study was to translate and culturally adapt the SOBBS to the Chinese context, specifically for young Chinese women, to address some limitations of the existing measures of self-objectification. To achieve this,

TABLE 1 | Descriptive statistics and correlations among items.

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Item11	Item12	Item13	Item14
Mean	2.98	2.37	3.23	2.26	3.03	2.05	2.87	3.08	2.96	2.07	2.35	2.05	2.70	2.05
SD	0.96	0.91	0.92	0.86	0.97	0.81	1.04	1.01	0.97	0.89	0.96	0.93	1.02	0.96
Kurtosis	−0.27	0.05	−0.07	0.61	−0.57	0.24	−0.79	−0.54	−0.56	0.23	0.02	0.75	−0.64	0.41
Skewness	−0.12	0.48	−0.40	0.65	−0.12	0.60	−0.10	−0.40	−0.08	0.68	0.51	0.88	−0.10	0.83
Item1	–													
Item2	0.42**	–												
Item3	0.41**	0.42**	–											
Item4	0.29**	0.53**	0.32**	–										
Item5	0.36**	0.35**	0.47**	0.37**	–									
Item6	0.28**	0.42**	0.27**	0.55**	0.31**	–								
Item7	0.40**	0.36**	0.49**	0.34**	0.47**	0.35**	–							
Item8	0.48**	0.42**	0.53**	0.37**	0.54**	0.38**	0.58**	–						
Item9	0.37**	0.38**	0.42**	0.38**	0.44**	0.31**	0.38**	0.50**	–					
Item10	0.28**	0.36**	0.24**	0.48**	0.29**	0.46**	0.30**	0.34**	0.35**	–				
Item11	0.32**	0.31**	0.23**	0.37**	0.31**	0.33**	0.27**	0.32**	0.33**	0.45**	–			
Item12	0.29**	0.40**	0.30**	0.43**	0.28**	0.54**	0.35**	0.35**	0.30**	0.52**	0.39**	–		
Item13	0.45**	0.41**	0.48**	0.38**	0.46**	0.45**	0.51**	0.58**	0.43**	0.42**	0.43**	0.58**	–	
Item14	0.26**	0.35**	0.24**	0.45**	0.24**	0.42**	0.33**	0.30**	0.27**	0.47**	0.41**	0.49**	0.43**	–

*SD, Standard deviation; **p < 0.01.*

we performed several steps. First, we translated the SOBBS into Chinese and obtained the Chinese version of the SOBBS. Unlike the SOQ (Fredrickson et al., 1998) and the FQSO (Wu and Lang, 2019) that use body attributes as items, the C-SOBBS utilizes unambiguous statements as items. These are more readily understood than body attributes, and they also avoid situations where body attributes do not capture group experiences. The Chinese version offers a new instrument for testing self-objectification in Chinese women to answer calls from Moradi (2010) and Calogero (2011) to clarify the construct and refine the assessment. We validated the two-factor structure of the C-SOBBS *via* CFA and evaluated the scale's construct and convergent validity, as well as incremental validity relative to the FQSO and OBC-Surveillance. Furthermore, we analyzed the internal consistency reliability and test-retest reliability of the C-SOBBS and its factors. The CFA results support the fit of the replicable two-factor model. Additionally, by analyzing the relationship between the overall C-SOBBS and its factors with other Chinese-language self-report measures of trait self-objectification, namely, the FQSO (Wu and Lang, 2019) and the OBC-Surveillance (McKinley and Hyde, 1996), as well as body shame assessed by the BSS of OBCS

TABLE 2 | Item standardized regression weights, squared multiple correlations (SMC), and item descriptive statistics for the C-SOBBS ($n = 331$).

Item	Weight	SMC
Observer's perspective		
1. I have thoughts about how my body looks to others even when I am alone	0.59	0.35
3. I try to imagine what my body looks like to others (i.e., like I am looking at myself from the outside)	0.67	0.44
5. I choose specific clothing or accessories based on how they make my body appear to others	0.66	0.43
7. When I look in the mirror, I notice areas of my appearance that I think others will view critically	0.70	0.49
8. I consider how my body will look to others in the clothing I am wearing	0.79	0.63
9. I often think about how my body must look to others	0.62	0.38
13. I try to anticipate others' reactions to my physical appearance	0.75	0.56
Body as self		
2. Looking attractive to others is more important to me than being happy with who I am inside	0.62	0.39
4. How I look is more important to me than how I think or feel	0.71	0.51
6. My physical appearance is more important than my personality	0.70	0.49
10. My physical appearance says more about who I am than my intellect	0.68	0.47
11. How sexually attractive others find me says something about who I am as a person	0.56	0.32
12. My physical appearance is more important than my physical abilities	0.71	0.50
14. My body is what gives me value to other people	0.64	0.41

TABLE 3 | Correlations among the measures.

Measures	Correlations						
	1	2	3	4	5	6	7
C-SOBBS_F1	–						
C-SOBBS_F2	0.64**	–					
C-SOBBS_Total	0.92**	0.89**	–				
FQSO	0.35**	0.50**	0.46**	–			
OBC-Surveillance	0.62**	0.51**	0.63**	0.49**	–		
BSS	0.32**	0.45**	0.40**	0.40**	0.38**	–	
ISOS	0.20**	0.29**	0.14**	0.14*	0.20**	0.13*	–
M	2.98	2.17	2.57	–0.86	3.85	3.43	22.35
SD	0.72	0.65	0.62	0.93	0.75	0.90	7.29
Possible scores	1 to 5	1 to 5	1 to 5	–6 to 6	1 to 15	1 to 7	1 to 15

*C-SOBBS_F1, Chinese version of Self-Objectification Beliefs and Behaviors-Observer's Perspective; C-SOBBS_F2, Chinese version of Self-Objectification Beliefs and Behaviors-Body as Self; C-SOBBS_Total, Chinese version of Self-Objectification Beliefs and Behaviors-Total Score; FQSO, Female Questionnaire of Trait Self-Objectification; OBC-Surveillance, the Self-Surveillance Subscale; BSS, Body Shame Subscale; and ISOS, Interpersonal Sexual Objectification Scale. **$p<0.01$; *$p<0.05$.*

TABLE 4 | Incremental validity of C-SOBBS relative to the FQSO and OBC-surveillance scale.

Outcome variable	ΔR^2	$\Delta R^2(F)$	*B*	*SE B*	β	*t*
BSS						
Model 1	0.16	62.37***				
FQSO			3.10	0.39	0.40	7.90***
Model 2	0.05	18.96***				
FQSO			2.17	0.44	0.28	4.96***
OBC-Surveillance			0.30	0.07	0.25	4.35***
Model 3	0.03	14.40***				
FQSO			1.79	0.44	0.23	4.05***
OBC-Surveillance			0.14	0.08	0.12	1.83
C-SOBBS			2.80	0.74	0.24	3.80***

*C-SOBBS, Chinese version of Self-Objectification Beliefs and Behaviors-Total Score; FQSO, the Female Questionnaire of Trait Self-Objectification; OBC-Surveillance, Objectified Body Consciousness Scale Self-Surveillance Subscale; BSS, Objectified Body Consciousness Scale Body Shame Subscale; and ISOS, Interpersonal Sexual Objectification Scale. ***$p<0.001$.*

(McKinley and Hyde, 1996) and sexual objectification measured by the ISOS (Kozee et al., 2007), the results demonstrate that the factor of observer's perspective, the factor of the "body as self," and the overall C-SOBBS were positively correlated with the FQSO, OBC-Surveillance, BSS, and ISOS. In addition, the correlation between the C-SOBBS and its factors is higher than the correlation between the C-SOBBS factor 1 (observer's perspective) and C-SOBBS factor 2 (body as self). These findings are consistent with the theoretical assumptions and provide evidence regarding the validity of the C-SOBBS as a suitable tool to measure self-objectification in young Chinese women. Additionally, the results of the hierarchical multiple regression demonstrate the incremental validity of the C-SOBBS. Furthermore, the stability of C-SOBBS by comparing two administrations of C-SOBBS over a two-week interval suggests that C-SOBBS scores are stable and indicative of a trait construct.

In general, the study presented in this paper provides evidence of a strong and replicable factor structure, construct, convergent, and incremental validity, as well as test-retest reliability for the Chinese version of the SOBBS.

Limitations and Future Research Directions

This research has several strengths, including the finding of evidence of reliability and validity for the Chinese version of the SOBBS. Nevertheless, several limitations of this study must be acknowledged.

First, the study relies exclusively on self-reported data, which may be susceptible to social desirability bias (Morgado et al., 2017), even though all the measures we use report high internal reliability. Second, our sample consists of female college students aged 18–35 years, with an average age of 20.28. The distribution of different age groups is inconsistent; the age group of 25–35 years is significantly under-represented compared to the 18–25 age group. Moreover, we did not recruit adolescent girls, older women, or men to participate in our study. Future studies need to further expand the sample size to verify the structure of C-SOBBS in older female groups as well as in adolescent girls and men.

Furthermore, the establishment of validation for any measure is an ongoing process. Although this study provides evidence of the reliability and validity of C-SOBBS among young Chinese women aged 18–35, it is still unclear whether its factor structure applies to other age groups. Future psychometric evaluations should investigate the issue of discriminant validity.

Practical Implications

The findings of this study have several important practical implications, including clinical ramifications. A substantial body of the literature shows that self-objectification is directly related to women's mental health. Women with a high level of self-objectification are associated with a high risk of physical anxiety (Tiggemann and Andrew, 2012; Watson et al., 2012), body dissatisfaction (Lindner et al., 2012; Tiggemann and Andrew, 2012; Brock et al., 2021), body shame (Tiggemann and Boundy, 2008; Choma et al., 2009; Schaefer et al., 2018; Baildon et al., 2021), depression (Peat and Muehlenkamp, 2011; Jones and Griffiths, 2015; Register et al., 2015), disordered eating (Schaefer and Thompson, 2018; Al-Mutawa et al., 2019; Kilpela et al., 2019; Holmes et al., 2020), and sexual dysfunction (Fredrickson and Roberts, 1997; Tiggemann, 2011). Counselors and therapists can use C-SOBBS to help people deal with issues related to self-objectification. Additionally, by focusing on attitudes and behaviors representative of self-objectification, C-SOBBS helps us identify potential intervention targets and can also serve to improve people's self-awareness of their tendencies to interpret how others may view their bodies while cultivating an appreciation for other positive aspects beyond physical appearance (Tylka and Augustus-Horvath, 2011; Calogero and Tylka, 2014; Lindner and Tantleff-Dunn, 2017).

CONCLUSION

In China, parallel to the increased prevalence of eating disorders and medical cosmetology, many researchers have conducted in-depth studies on self-objectification. However, the absence of a reliable and validated measurement tool specifically constructed to assess self-objectification is a serious impediment to accurate and high-quality research. The purpose and primary achievement of this article is to address this gap by validating an additional instrument adapted specifically to measure young Chinese women's self-objectification. In this study, we provide sufficient reliability and validity for C-SOBBS, concluding that it is an effective tool with the potential to foster further research that will advance related theories, research, and practices.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Fujian Normal University Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ML: conceptualization, formal analysis, and writing. YY: conceptualization, editing, and supervision. All authors contributed to the article and approved the submitted version.

REFERENCES

Adams, K. E., Tyler, J. M., Calogero, R., and Lee, J. (2017). Exploring the relationship between appearance-contingent self-worth and self-esteem: The roles of self-objectification and appearance anxiety. *Body Image* 23, 176–182. doi: 10.1016/j.bodyim.2017.10.004

Al-Mutawa, N., Schuilenberg, S. J., Justine, R., and Taher, S. K. (2019). Modesty, objectification, and disordered eating patterns: a comparative study between veiled and unveiled Muslim women residing in Kuwait. *Med. Princ. Pract.* 28, 41–47. doi: 10.1159/000495567

Augustus-Horvath, C. L., and Tylka, T. L. (2009). A test and extension of objectification theory as it predicts disordered eating: does women's age matter? *J. Couns. Psychol.* 56, 253–265. doi: 10.1037/a0014637

Baildon, A. E., Eagan, S. R., Christ, C. C., Lorenz, T., Stoltenberg, S. F., and Gervais, S. J. (2021). The sexual objectification and alcohol use link: the mediating roles of self-objectification, enjoyment of sexualization, body shame, and drinking motives. *Sex Roles* 85, 190–204. doi: 10.1007/s11199-020-01213-2

Bartky, S. L. (1990). *Femininity and Domination: Studies in the Phenomenology of Oppression.* New York: Routledge.

Beaton, D. E., Bombardier, C., Guillemin, F., and Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25, 3186–3191. doi: 10.1097/00007632-200012150-00014

Brock, R. L., Ramsdell, E. L., Saez, G., and Gervais, S. J. (2021). Perceived humanization by intimate partners during pregnancy is associated with fewer depressive symptoms, less body dissatisfaction, and greater sexual satisfaction through reduced self-objectification. *Sex Roles* 84, 285–298. doi: 10.1007/s11199-020-01166-6

Calogero, R. M. (2011). "Chapter 2. Operationalizing self-objectification: assessment and related methodological issues" in *Self-Objectification in Women: Cause, Consequences, and Counteractions*. eds. R. M. Calogero, S. Tantleff-Dunn and J. K. Thompson (Washington: American Psychological Association), 23–49.

Calogero, R. M., and Tylka, T. L. (2014). Sanctioning resistance to sexual objectification: an integrative system justification perspective. *J. Soc. Issues* 70, 763–778. doi: 10.1111/josi.12090

Chen, X., and Jiang, Y. J. (2007). The revision of body consciousness scale for college students. *Chin. Ment. Health J.* 21, 610–613. doi: 10.3321/j.issn:1000-6729.2007.09.009

Choma, B. L., Shove, C., Busseri, M. A., Sadava, S. W., and Hosker, A. (2009). Assessing the role of body image coping strategies as mediators or moderators of the links between self-objectification, body shame, and well-being. *Sex Roles* 61, 699–713. doi: 10.1007/s11199-009-9666-9

Cohen, R., Newton-John, T., and Slater, A. (2018). 'Selfie'-objectification: the role of selfies in self-objectification and disordered eating in young women. *Comput. Hum. Behav.* 79, 68–74. doi: 10.1016/j.chb.2017.10.027

Fleiss, J. L. (1999). *The Design and Analysis of Clinical Experiments*. New York, NY: Wiley.

Fornell, C., and Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18, 39–50. doi: 10.1177/002224378101800104

Fredrickson, B. L., and Roberts, T. A. (1997). Objectification theory. *Psychol. Women Q.* 21, 173–206. doi: 10.1111/j.1471-6402.1997.tb00108.x

Fredrickson, B. L., Roberts, T. A., Noll, S. M., Quinn, D. M., and Twenge, J. M. (1998). That swimsuit becomes you: sex differences in self-objectification, restrained eating, and math performance. *J. Pers. Soc. Psychol.* 75, 269–284. doi: 10.1037/0022-3514.75.1.269

Grippo, K. P., and Hill, M. S. (2008). Self-objectification, habitual body monitoring, and body dissatisfaction in older European American women: exploring age and feminism as moderators. *Body Image* 5, 173–182. doi: 10.1016/j.bodyim.2007.11.003

Hill, M. S., and Fischer, A. R. (2008). Examining objectification theory: lesbian and heterosexual women's experiences with sexual- and self-objectification. *Couns. Psychol.* 36, 745–776. doi: 10.1177/0011000007301669

Holmes, S. C., DaFonseca, A. M., and Johnson, D. M. (2020). Sexual victimization and disordered eating in bisexual women: a test of objectification theory. *Violence Against Women* 27, 2021–2042. doi: 10.1177/1077801220963902

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Jones, B. A., and Griffiths, K. M. (2015). Self-objectification and depression: an integrative systematic review. *J. Affect. Disord.* 171, 22–32. doi: 10.1016/j.jad.2014.09.011

Jongenelis, M. I., Byrne, S. M., and Pettigrew, S. (2014). Self-objectification, body image disturbance, and eating disorder symptoms in young Australian children. *Body Image* 11, 290–302. doi: 10.1016/j.bodyim.2014.04.002

Kilpela, L. S., Calogero, R., Wilfred, S. A., Verzijl, C. L., Hale, W. J., and Becker, C. B. (2019). Self-objectification and eating disorder pathology in an ethnically diverse sample of adult women: cross-sectional and short-term longitudinal associations. *J. Eat. Disord.* 7:45. doi: 10.1186/s40337-019-0273-z

Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling*. New York: Guilford Press.

Kozee, H. B., Tylka, T. L., Augustus-Horvath, C. L., and Denchik, A. (2007). Development and psychometric evaluation of the Interpersonal Sexual Objectification Scale. *Psychol. Women Q.* 31, 176–189. doi: 10.1111/j.1471-6402.2007.00351.x

Lindner, D., and Tantleff-Dunn, S. (2017). The development and psychometric evaluation of the Self-Objectification Beliefs and Behaviors Scale. *Psychol. Women Q.* 41, 254–272. doi: 10.1177/0361684317692109

Lindner, D., Tantleff-Dunn, S., and Jentsch, F. (2012). Social comparison and the 'circle of objectification'. *Sex Roles* 67, 222–235. doi: 10.1007/s11199-012-0175-x

MacCallum, R., Browne, M., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychol. Methods* 1, 130–149. doi: 10.1037/1082-989X.1.2.130

McKinley, N. M., and Hyde, J. S. (1996). The Objectified Body Consciousness Scale development and validation. *Psychol. Women Q.* 20, 181–215. doi: 10.1111/j.1471-6402.1996.tb00467.x

Moradi, B. (2010). Addressing gender and cultural diversity in body image: objectification theory as a framework for integrating theories and grounding research. *Sex Roles* 63, 138–148. doi: 10.1007/s11199-010-9824-0

Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., and Ferreira, M. E. C. (2017). Scale development: ten main limitations and recommendations to improve future research practices. *Psicol. Reflex. Crit.* 30:3. doi: 10.1186/s41155-017-0059-7

Nunnally, J. C. (1978). *Psychometric Theory*. 2nd Edn. New York: McGraw-Hill.

Peat, C. M., and Muehlenkamp, J. J. (2011). Self-objectification, disordered eating, and depression: a test of mediational pathways. *Psychol. Women Q.* 35, 441–450. doi: 10.1177/0361684311400389

Register, J. D., Katrevich, A. V., Aruguete, M. S., and Edman, J. L. (2015). Effects of self-objectification on self-reported eating pathology and depression. *Am. J. Psychol.* 128, 107–113. doi: 10.5406/amerjpsyc.128.1.0107

Schaefer, L. M., Burke, N. L., Calogero, R. M., Menzel, J. E., Krawczyk, R., and Thompson, J. K. (2018). Self-objectification, body shame, and disordered eating: Testing a core mediational model of objectification theory among White, Black, and Hispanic women. *Body Image* 24, 5–12. doi: 10.1016/j.bodyim.2017.10.005

Schaefer, L. M., and Thompson, J. K. (2018). Self-objectification and disordered eating: A meta-analysis. *Int. J. Eat. Disord.* 51, 483–502. doi: 10.1002/eat.22854

Sun, Q. Q., and Zheng, Y. (2016). Sexual objectification and self-objectification: investigating the relationship and the underlying neurocognitive. unpublished doctoral dissertation. Chongqing: Southwestern University.

Swami, V., and Barron, D. (2019). Translation and validation of body image instruments: Challenges, good practice guidelines, and reporting recommendations for test adaptation. *Body Image* 31, 204–220. doi: 10.1016/j.bodyim.2018.08.014

Talmon, A., and Ginzburg, K. (2016). The nullifying experience of self-objectification: The development and psychometric evaluation of the Self-Objectification Scale. *Child Abuse Negl.* 60, 46–57. doi: 10.1016/j.chiabu.2016.09.007

Teng, F., Gao, W. Y., Huang, X. S., and Poon, K. T. (2019). Body surveillance predicts men's and women's perceived loneliness: a serial mediation model. *Sex Roles* 81, 97–108. doi: 10.1007/s11199-018-0977-6

Tiggemann, M. (2011). "Mental health risks of self-objectification: a review of the empirical evidence for disordered eating, depression mood, and sexual dysfunction" in *Self-Objectification in Women: Cause, Consequences, and Counteractions*. eds. R. M. Calogero, S. Tantleff-Dunn and J. K. Thompson (Washington: American Psychological Association), 139–155.

Tiggemann, M., and Andrew, R. (2012). Clothes make a difference: the role of self-objectification. *Sex Roles* 66, 646–654. doi: 10.1007/s11199-011-0085-3

Tiggemann, M., and Boundy, M. (2008). Effect of environment and appearance compliment on college women's self-objectification, mood, body shame, and cognitive performance. *Psychol. Women Q.* 32, 399–405. doi: 10.1111/j.1471-6402.2008.00453.x

Tiggemann, M., and Kuring, J. K. (2004). The role of body objectification in disordered eating and depressed mood. *Br. J. Clin. Psychol.* 43, 299–311. doi: 10.1348/0144665031752925

Tylka, T. L., and Augustus-Horvath, C. L. (2011). "Fighting self-objectification in prevention and intervention contexts" in *Self-Objectification in Women: Causes, Consequences, and Counteractions*. eds. R. M. Calogero, S. Tantleff-Dunn and J. K. Thompson (Washington: American Psychological Association), 187–214.

Vencill, J. A., Tebbe, E. A., and Garos, S. (2015). It's not the size of the boat or the motion of the ocean: the role of self-objectification, appearance anxiety, and depression in female sexual functioning. *Psychol. Women Q.* 39, 471–483. doi: 10.1177/0361684315587703

Wang, Y. H., Wang, X. C., Yang, J., Zeng, P., and Lei, L. (2020). Body talk on social networking sites, body surveillance, and body shame among young adults: the roles of self-compassion and gender. *Sex Roles* 82, 731–742. doi: 10.1007/s11199-019-01084-2

Watson, L. B., Robinson, D., Dispenza, F., and Nazari, N. (2012). African American women's sexual objectification experiences: a qualitative study. *Psychol. Women Q.* 36, 458–475. doi: 10.1177/0361684312454724

Wu, M. X., and Lang, M. (2019). Female Questionnaire of Trait Self-Objectification: initial development and validation in China. *Sex Roles* 80, 758–769. doi: 10.1007/s11199-018-0972-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 13 October 2021
doi: 10.3389/fpsyg.2021.754236

Validation of a Kazakhstani Version of the Mental Health Continuum—Short Form

Daniel Hernández-Torrano [1], Laura Ibrayeva [1], Ainur Muratkyzy [1], Natalya Lim [2], Yerden Nurtayev [3], Ainur Almukhambetova [1], Alessandra Clementi [2] and Jason Sparks [1]*

[1] *Graduate School of Education, Nazarbayev University, Nur-Sultan, Kazakhstan,* [2] *Department of Medicine, Nazarbayev University School of Medicine, Nur-Sultan, Kazakhstan,* [3] *Health and Wellness Center, Nazarbayev University, Nur-Sultan, Kazakhstan*

Positive mental health and well-being are significant dimensions of health, employment, and educational outcomes. Research on positive mental health and well-being requires measurement instruments in native languages for use in local contexts and target populations. This study examines the psychometric properties of the Kazakhstani version of the Mental Health Continuum—Short Form (MHC-SF), a brief self-report instrument measuring emotional, social, and psychological well-being. The sample included 664 University students (425 females) purposefully selected in three higher education institutions in South, East, and Central Kazakhstan. Their average age was 20.25 and ranged from 18 to 43. Participants completed a Kazakhstani version of the MHC-SF online. Statistical analyses to evaluate the structural validity, reliability, and measurement invariance of the Kazakhstani version of the MHC-SF were performed. The results confirmed the superiority of the bifactor model (i.e., three separated factors of well-being plus a general factor of well-being) over the alternatives. However, most of the reliable variance was attributable to the general well-being factor. Subscale scores were unreliable, explaining very low variance beyond that explained by the general factor. The findings demonstrated the measurement invariance of the MHC-SF across gender and age. Overall, these findings support the use of the Kazakhstani version of the MHC-SF to examine a general factor of well-being and the measurement invariance of the instrument across gender and age groups. However, the results advise against the interpretation of the subscale scores as unequivocal indicators of emotional, social, and psychological well-being.

Keywords: mental health, well-being, MHC-SF, health psychology, validation, psychometric adequacy, Kazakhstan, Central Asia

OPEN ACCESS

Edited by:
Victor Zaia,
Faculdade de Medicina do ABC, Brazil

Reviewed by:
Martina Rahe,
University of Koblenz-Landau, Germany
Richard Appiah,
University of Ghana, Ghana

****Correspondence:***
Daniel Hernández-Torrano
daniel.torrano@nu.edu.kz;
d.hernandeztorrano@gmail.com

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *06 August 2021*
Accepted: *15 September 2021*
Published: *13 October 2021*

Citation:
Hernández-Torrano D, Ibrayeva L, Muratkyzy A, Lim N, Nurtayev Y, Almukhambetova A, Clementi A and Sparks J (2021) Validation of a Kazakhstani Version of the Mental Health Continuum—Short Form. Front. Psychol. 12:754236. doi: 10.3389/fpsyg.2021.754236

INTRODUCTION

The WHO defines mental health as "a state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community" (World Health Organization, 2004, p. 12). This definition breaks with a tradition based on psychopathological-oriented models of mental health with a focus on disorders and illness toward a conceptualization that pays more attention to the presence of positive features and what is right about people (Kobau et al., 2011).

It builds upon two long-standing traditions in positive psychology: hedonic and eudaimonic well-being. The hedonic approach is connected to happiness and defines well-being as the presence of positive feelings and pleasure and the absence of negative feelings or pain. The eudaimonic approach focuses on self-realization and meaning and defines well-being as the degree to which a person is functioning appropriately (Ryan and Deci, 2001; Keyes, 2006, 2007; Jovanović, 2015; Joshanloo and Lamers, 2016).

International interest in positive mental health and well-being has grown exponentially in recent decades due to the significant effects they have on health, employment, and educational outcomes (see Keyes, 2013, for an international review of correlates to mental well-being in these domains). Research on positive mental health and well-being requires measurement instruments that provide reliable and valid scores in different contexts, cultures, and languages. The Mental Health Continuum—Short Form (MHC-SF) (Keyes et al., 2008) is one of the most widely used self-report instruments to measure positive mental health and well-being around the world in clinical (e.g., Silverman et al., 2018; van Erp Taalman Kip and Hutschemaekers, 2018; Donnelly et al., 2019), work (e.g., Jaotombo, 2019), and educational settings (e.g., de Carvalho et al., 2016; Luijten et al., 2019).

The MHC-SF is a 14-item measure of positive mental health that encompasses both hedonic and eudaimonic well-being (Keyes et al., 2008). It was designed to measure three dimensions of positive mental health: (1) emotional well-being, which refers to the presence of positive feelings and life satisfaction; (2) social well-being, which accounts for adequate social functioning and connection to society; and (3) psychological well-being, which reflects personal functioning and thriving in life. In the MHC-SF, emotional well-being, social well-being and psychological well-being are measured using three, five, and six items, respectively. Emotional well-being reflects the hedonic approach to well-being, whereas social and psychological well-being are used as indicators of eudaimonic well-being (see Lamers et al., 2011; Petrillo et al., 2015).

Along with these three dimensions of well-being, the MHC-SF can also help categorize three states of mental health: flourishing, moderate mental health, and languishing (Keyes, 2002). Flourishing represents high levels of well-being in both hedonic and eudaimonic well-being. Languishing in a mental health state characterized by lower scores in hedonic and eudaimonic well-being. Individuals who do not meet the criteria for either flourishing and languishing are considered to demonstrate moderate levels of mental health (Keyes, 2005). The results of the first study using the MHC-SF demonstrated that 12.2% of the participants were languishing, 67.8% were moderately mentally healthy, and 20% were flourishing (Keyes et al., 2008). Other studies with adult populations have found similar distributions of the categorical diagnosis of states of mental health, with 10–20% of participants languishing, 50–70% moderately mentally healthy, and 20–30% flourishing (e.g., Lamers et al., 2011; Petrillo et al., 2015), although there is wide variability across contexts. Younger samples such as adolescents and college students tend to demonstrate comparatively higher distributions of flourishing and moderately mentally healthy (e.g., Luijten et al., 2019; Hides et al., 2020).

The psychometric characteristics of the MHC-SF have been widely explored across various contexts, cultures, and languages, predominantly in Europe (e.g., Lamers et al., 2011; Karaś et al., 2014; Jovanović, 2015; Petrillo et al., 2015; Joshanloo and Lamers, 2016; Echeverría et al., 2017; Joshanloo and Jovanović, 2017; Donnelly et al., 2019; Luijten et al., 2019; Santini et al., 2020; Monteiro et al., 2021). Evidence about the appropriate reliability and validity of translated versions of the scale also exists in other continents, including Asia (Lim, 2014; Guo et al., 2015; Rafiey et al., 2017; Rogoza et al., 2018; Joshanloo, 2020), North America (Joshanloo, 2016a, 2018, 2019; Lamborn et al., 2018), South America (Contreras et al., 2017; Perugini et al., 2017), Africa (Keyes et al., 2008; De Bruin and Du Plessis, 2015), and Oceania (Hides et al., 2016; Joshanloo et al., 2017). Such studies are paramount because measuring instruments that have been developed and normed in one context (e.g., Western; more individualistically socially oriented) must be translated and validated before they can be administered in other contexts (e.g., Asian/African; more collectivistically socially oriented). This is especially true in the field of mental health and well-being because the issues of the "mind" are often interpreted very differently across contexts and countries (Fernando, 2019). Indeed, cultural values, traditions, and languages influence conceptions of mental health and well-being (Eshun and Gurung, 2009; Vaillant, 2012). No previous study has examined the properties of this instrument in Central Asia. This study aimed to fill this gap by examining the reliability and structural validity of the Kazakhstani version of the MHC-SF.

The factor structure of the MHC-SF has been extensively analyzed using exploratory factor analysis (EFA) (Lamers et al., 2011; Rafiey et al., 2017), confirmatory factor analysis (CFA; Karaś et al., 2014; Guo et al., 2015; Petrillo et al., 2015), and exploratory structural equation modeling (ESEM) (Joshanloo and Lamers, 2016; Joshanloo and Jovanović, 2017). Joshanloo (2018) has also recently examined the internal structure of the MHC-SF using a multidimensional scaling approach. Overall, the results of these studies support the original structure of the questionnaire, with three-factor models (i.e., emotional, social, and psychological well-being) demonstrating a better fit than one-factor (i.e., mental health), two-factor (i.e., hedonic and eudaimonic well-being), and second-order factor models (i.e., three correlated first-order factors representing emotional, social, and psychological well-being loading into a second-order general well-being factor).

However, there is a growing concern regarding the goodness of fit of the three-factor model. An alternative bifactor model has been recently proposed as a superior explanation of the internal structure of the MHC-SF (e.g., De Bruin and Du Plessis, 2015; Jovanović, 2015; Echeverría et al., 2017; Schutte and Wissing, 2017; Rogoza et al., 2018; Silverman et al., 2018; Longo et al., 2020). The bifactor model accounts for a general well-being factor and three separate well-being factors capturing specific variance for emotional, social, and psychological well-being. In general, these studies argue that the bifactor model fits the data better than alternative models. In the bifactor model, the general well-being factor accounts for a greater amount of variance than the three specific well-being factors, and after controlling for the variance of the general factor, the three specific factors explain a very little amount of variance. Moreover, a strong general factor of well-being seems to emerge in these studies, no matter whether a CFA bifactor or an ESEM bifactor approach are used (Longo et al., 2020).

Studies exploring the three-factor solution of the MHC-SF consistently report satisfactory internal consistency (Keyes et al., 2008; Lamers et al., 2011; Perugini et al., 2017; Luijten et al., 2019) and test-retest reliability (e.g., Petrillo et al., 2015). Cronbach's coefficient α for the general and specific subscale scores in the bifactor model seem to be also appropriate across studies (e.g., De Bruin and Du Plessis, 2015; Echeverría et al., 2017). However, when alternative reliability coefficients such as McDonald's ω are used to estimate the reliability of the scale scores in the bifactor model, the findings are less conclusive. In some cases, coefficients ω suggest good reliability of the scores for the general and separate well-being subscales (e.g., Lamborn et al., 2018; Rogoza et al., 2018). In other cases, findings suggest that the general factor tends to account for a greater amount of variance of the MHC-SF and that three subscales demonstrate low reliability and explain very little variance beyond that explained by the general factor (Jovanović, 2015; Echeverría et al., 2017; Schutte and Wissing, 2017; Silverman et al., 2018; Longo et al., 2020; Santini et al., 2020).

The measurement invariance of the MHC-SF has also been explored in the literature. In general, there is considerable evidence supporting the invariance of the structure of the MHC-SF across different groups, contexts, and conditions. Measurement invariance across gender and age has been observed in most studies for both the three-factor (e.g., Karaś et al., 2014; Guo et al., 2015; Joshanloo and Jovanović, 2017; Joshanloo et al., 2017; Perugini et al., 2017) and bifactor solutions (e.g., Echeverría et al., 2017; Lamborn et al., 2018) of the MHC-SF. The longitudinal measurement invariance of the bifactor model has also been examined, with no apparent differential item functioning over time (Lamers et al., 2011). Moreover, there is growing evidence for the full or partial cross-cultural measurement invariance of the instrument. For example, Joshanloo (2016a) demonstrated that the items of the MHC-SF function relatively similarly across samples in Iran and the USA when considering the three-factor model. Schutte and Wissing (2017) also reported evidence for the partial invariance of the bifactor model across three cultural groups in South Africa. Similarly, Zemojtel-Piotrowska et al. (2018) investigated the cross-cultural measurement invariance of the MHC-SF across 38 countries, confirming "the cross-cultural replicability of a bifactor structure" (p. 1035).

MATERIALS AND METHODS

Participants

The sample included 664 University students purposefully selected in three higher education institutions located in South, East, and Central Kazakhstan. Among them, 425 were females (64.0%), 236 were males (35.5%), and the 3 did not report their gender (0.5%). Their age ranged from 18 to 43 ($M = 20.25$, $SD = 3.61$). A total of 383 were under 20 years old (57.7%), and 281 were 20 years-old or older (42.3%). Most of the participants were single and had no children. A majority of them were studying an undergraduate program (79.7%). The rest were enrolled in a master's program (11.9%), a PhD program (3.2%), or other programs (5.2%). From the total sample, 11% were studying a major in Natural Sciences, 49% in Technical Sciences, 30% in Social Sciences, and 10% in Humanities.

Instrument and Procedures

The Mental Health Continuum-Short Form (MHC-SF) (Keyes et al., 2008) is a brief questionnaire that measures positive mental health. It comprises 14 items that represent several feelings of well-being: three items reflect emotional well-being (items 1–3) (e.g., In the past month, how often did you feel happy), five items reflect social well-being (items 4–8) (e.g., In the past month, how often did you feel that you had something important to contribute to society), and six items reflect psychological well-being (items 9–14) (e.g., In the past month, how often did you feel that you liked most parts of your personality). Participants rate the frequency of each feeling in the past month on a 6-point Likert scale (0 = never, 1 = once or twice, 2 = about once a week, 3 = about 2 or 3 times a week, 4= almost every day, 5 = every day). The MHC-SF offers two levels of assessment. First, it allows for the evaluation of the three well-being dimensions (i.e., emotional, social, and psychological). Second, a categorical diagnosis of mental health status with three categories: flourishing, moderate, and languishing.

The MHC-SF was translated into the two official languages of Kazakhstan (i.e., Russian and Kazakh) using a back-translation approach (Brislin, 1970). Two members of the research team who were native speakers translated the MHC-SF into Russian and Kazakh languages. Next, independent translators who were unfamiliar with the original version of the instrument translated these versions back to English language. The research team then examined the original and back translated versions to ensure comparability. In addition to that, the Russian and Kazakh translations of the MHC-SF were further assessed by the research team to ensure understandability, psychological equivalence, and the accuracy of the translations (Douglas and Craig, 2007).

The Kazakhstani version of the MHC-SF was distributed online via email by the gatekeepers of the respective universities. Participants provided informed consent before proceeding to complete the questionnaire. Anonymity and confidentiality were ensured and no information that could identify the identities of the participants was collected. The study was approved by the Ethics Committee of the authors' institution (reference number 195/19112019).

Statistical Analysis

Descriptive statistics were calculated to examine the distribution of the scores. Relevant subscales items were summed to yield a score for emotional well-being (items 1–3; range 0–15), social well-being (items 4–8; range 0–25), and psychological well-being (items 9–14; range 0–30). A total well-being score was obtained by summing up the 14 items of the scale (range 0–70).

The categorical diagnosis using the MHC-SF by Keyes (2006) was applied to the data to obtain estimates of the population prevalence of the mental health categories. A diagnosis of flourishing is made if someone feels one of the three hedonic well-being symptoms (items 1–3) "every day" or "almost every day" and feels six of the 11 positive functioning symptoms (items 4–14) "every day" or "almost every day" in the past month. A person is diagnosed as languishing if they feel the three hedonic well-being symptoms "never" or "once or twice" and six of the 11 eudaimonic well-being symptoms "never" or "once or twice" in the past month. Individuals not meeting neither "languishing" nor "flourishing" criteria are diagnosed as "moderately mentally healthy."

Confirmatory Factor Analyses (CFA) were conducted to examine the structural validity of the Kazakhstani version of the MHC-SF with the *lavaan* (Rosseel, 2012) and *semPlot* (Epskamp, 2015) packages in R (R Core Team, 2020). Based on the theory and previous research (see Introduction section), five distinctive models were tested in this study: (1) a single factor model in which all the 14 items load into one general factor of well-being; (2) a two-factor model with two correlated factors: hedonic well-being (EWB, items 1–3), and eudaimonic well-being (SWB and PWB, items 4–14); (3) a three-factor model, with three correlated factors of well-being (EWB, items 1–3; SWB, items 4–8; PWB, items 9–14 PWB); (4) a second-order model were the three first order dimensions (EWB, SWB, PWB) load in a general factor of well-being; and (5) a bifactor model, with three separated factors of well-being (EWB, SWB, PWB) plus a general factor of well-being.

Participants with one or more missing data were excluded from the analysis. The Satorra-Bentler scaled chi-square test (SB χ^2) test was used to evaluate the absolute fit of the model. However, because the SB χ^2 test is considered highly conservative, additional absolute and incremental alternative fit indices were used to evaluate the model. Additional absolute fit indices included the SB χ^2 to degrees of freedom ratio (SB χ^2/df), the Root Mean Square Error of Approximation (RMSEA), the Standardized Root Mean Square Residual (SRMR), and the Akaike's Information Criterion (AIC). Indices of incremental fit comprised the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI). Values of SB $\chi^2/df < 3$, RMSEA and SRMR < 0.06, and CFI and TLI > 0.95 indicated a good model fit, while SB $\chi^2/df < 5$, RMSEA and SRMR < 0.08, and CFI and TLI > 0.90 indicated a satisfactory fit (Hu and Bentler, 1999; Schreiber et al., 2006). In general, lower AIC values indicate a better fit.

The reliability of the factors of the best fitting model was estimated using omega (ω) coefficients using the *psych* package in R (Revelle, 2021). Omega has demonstrated to be superior to Cronbach's alpha and other reliability coefficients to capture the proportion of scale variance due to all common factors and the proportion of scale variance due to a general factor (Zinbarg et al., 2005, p. 132), as is the case for the multidimensional models tested in this study. Omega reliability coefficients (ω) were calculated to estimate the proportion of variance in the observed total score attributable to both general and specific well-being factors as suggested by Rodriguez et al. (2016). Furthermore, omega hierarchical (ω_h) was calculated to estimate

TABLE 1 | Descriptive statistics.

	M	SD	Skewness	Kurtosis	Shapiro-Wilks
EWB	9.02	4.00	−0.40	−0.71	0.96***
SWB	10.78	6.29	0.20	−0.83	0.96***
PWB	17.45	7.80	−0.26	−0.92	0.97***
Total	37.03	16.26	−0.09	−0.82	0.98***
Mental health categories	Total	Female	Male	Younger	Older
Languishing	17.6%	18.1%	17.0%	16.2%	19.6%
Moderate	45.2%	48.0%	39.8%	45.4%	44.8%
Flourishing	27.9%	24.7%	33.5%	29.2%	25.9%

EWB, emotional well-being; SWB, social well-being; PSW, psychological well-being.
****p < 0.001.*

the proportion of variance in total scores attributable to a single general well-being factor (Rodriguez et al., 2016). Additionally, omega subscale (ω_s) was used to calculate the reliability of the subscale scores and hierarchical subscale (ω_{hs}) coefficients were used to estimate the amount of variance in each subscale that is explained by a specific factor (i.e., EWB, SWB, and PSW), after removing the reliable variance explained by the general well-being factor. Following Perreira et al. (2018), ω coefficients > 0.50 were considered satisfactory.

Gender and age invariance of the best-fitting factor model was examined using multigroup confirmatory factor analyses (MGCFA). We tested configural invariance, metric invariance, scalar invariance, and strict invariance across gender (male vs. female) and age (younger vs. older). Configural invariance was confirmed if RSMEA and SRMR were <0.08 and CFI was >0.95 (Cheung and Rensvold, 2002). A relative change of ≤0.010 in CFI, paired with a relative change of ≤0.015 in RMSEA and ≤0.030 in SRMR (for metric invariance) or ≤0.015 (for scalar and residual invariance) indicated support for measurement invariance (Cheung and Rensvold, 2002; Chen, 2007; Putnick and Bornstein, 2016).

RESULTS

Descriptive Analysis

Descriptive statistics along with the percentage of participants meeting the criteria for the three categorical diagnoses of mental health in the Kazakhstani version of the MHC-SF are presented in **Table 1**. Skewness and kurtosis did not exceed 1 for any subscale. However, the Shapiro-Wilk test was statistically significant for all subscales, revealing that the data were assumed to be not normally distributed. Based on the diagnostic criteria of mental health as measured by the MHC-SF, the findings revealed that 19.4% of the sample were languishing, 49.8% were moderately mentally healthy, and 30.7% were flourishing. Relevant differences were found in the gender distribution for the categorical diagnosis of languishing, $\chi^2 = 6.56$, $p = 0.038$, with males (33.5%) demonstrating higher distributions for flourishing than females (24.7%). No statistically significant differences were observed with respect to age, $\chi^2 = 1.68$, $p = 0.431$, although a slightly lower proportion of younger students (under 20 years

TABLE 2 | Confirmatory factor analysis fit statistics.

Model	SB χ^2 $_{(df)}$	SB χ^2 / *df*	CFI	TLI	RMSEA (90% CI)	SRMR	AIC
Single factor	615.35* $_{(77)}$	7.99	0.864	0.840	0.123 (0.114–0.132)	0.061	29,119.56
Two-factor	487.82* $_{(76)}$	6.41	0.898	0.878	0.107 (0.098–0.117)	0.056	28,932.64
Three-factor	303.30* $_{(74)}$	4.10	0.944	0.931	0.081 (0.071–0.090)	0.049	28,677.14
Second-order factor	303.30* $_{(74)}$	4.10	0.944	0.931	0.081 (0.071–0.090)	0.049	28,677.14
Bifactor	199.64* $_{(63)}$	3.16	0.967	0.953	0.067 (0.057–0.077)	0.030	28,554.91

SB χ^2, Satorra-Bentler scaled chi-square; df, degrees of freedom; CFI, comparative fit index; TLI, Tucker-Lewis Index; RMSEA, root square error of approximation; CI, confidence intervals; SRMR, standardized root mean square residual; AIC, Akaike's information criterion.
**p < 0.001.*

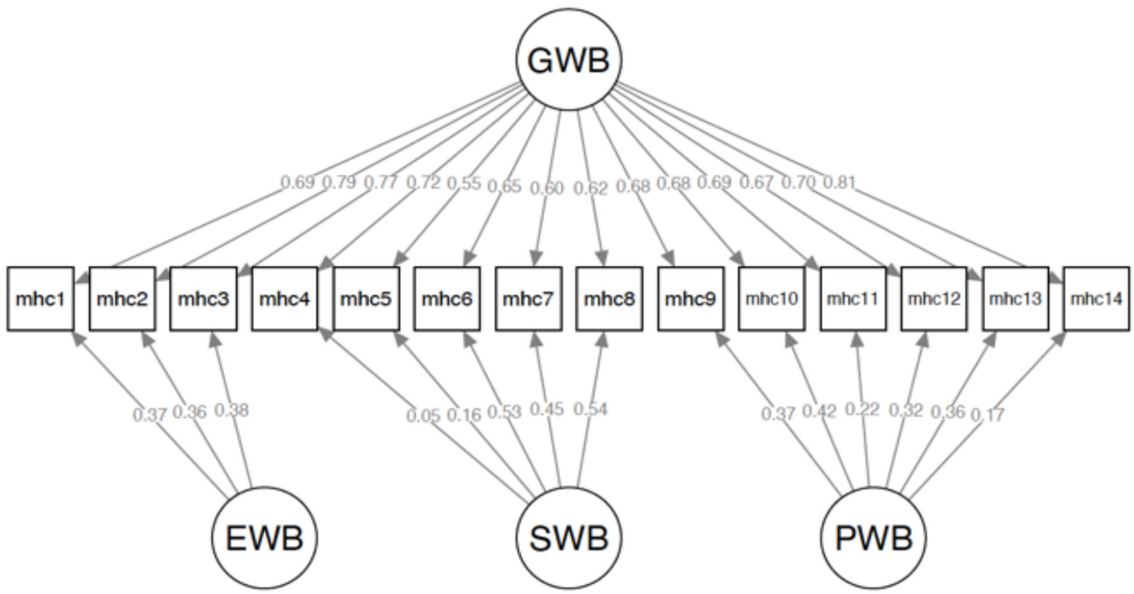

FIGURE 1 | Standardized factor loadings of the bifactor model. GWB, general well-being; EWB, emotional well-being; SWB, social well-being; PSW, psychological well-being.

old) classified as languishing (16.2 vs. 19.6%) and a higher proportion as flourishing (29.2 vs. 25.9%) when compared to older students (20 years or older).

Factor Structure of the MHC-SF

Confirmatory Factor Analyses (CFA) were conducted to examine the structural validity of the Kazakhstani version of the MHC-SF. The parameter estimates in the CFAs were obtained using the robust maximum likelihood (MLM) method to account for any deviations from normality (Brown, 2006). **Table 2** presents the CFA fit indices for the five models tested in this study. As indicated by the SB χ^2 values, none of the models fit perfectly. The single factor and the correlated two-factor models were found to have an absolute poor fit. The correlated three-factor and second-order factor models achieved satisfactory to good fit, with identical fitting indices as they are mathematically equivalent. The bifactor model demonstrated an overall good fit, with SB $\chi^2/df = 3.16$, CFI and TLI > 0.95, RMSEA and SRMR < 0.06, and the lowest AIC.

Substantial differences in SB χ^2/df, TLI, CFI, RMSEA, and SRMR between the bifactor model and alternative models demonstrated the superiority of the bifactor model. **Figure 1** presents the standardized path estimates for the bifactor model. All standardized path estimates loaded significantly in the general well-being factor ($\beta = 0.55$–0.81, $p <. 001$). Also, all items loaded significantly ($p < 0.01$) in the hypothesized specific factor, except for the item 4 ($\beta = 0.05$, $p = 0.272$). Noteworthy, all items exhibited higher loadings on the general factor than their respective specific factor (EWB, SWB, and PWB), indicating that the variances of the items were generally explained by the general well-being factor.

Internal Consistency

The omega reliability coefficient indicated that 93% of the variance was explained by both the general and three specific wellbeing factors ($\omega = 0.93$). The omega hierarchical coefficient (ω_h) was 0.86, suggesting that 86% of the variance of uni-weighted total scores could be explained by individual differences

TABLE 3 | Measurement invariance of the bifactor model across gender and age.

Model	SB χ^2 $_{(df)}$	CFI	RMSEA	ΔCFI	ΔRMSEA
Gender invariance					
Configural	319.11 $_{(126)}$	0.965	0.069	–	–
Metric	347.32 $_{(150)}$	0.964	0.064	0.001	0.005
Scalar	377.08 $_{(160)}$	0.961	0.065	0.004	0.001
Strict	390.26 $_{(164)}$	0.959	0.066	0.002	0.001
Age invariance					
Configural	357.71 $_{(126)}$	0.958	0.076	–	–
Metric	400.64 $_{(150)}$	0.955	0.072	0.003	0.004
Scalar	419.15 $_{(160)}$	0.953	0.071	0.002	0.001
Strict	431.46 $_{(164)}$	0.952	0.071	0.001	0.000

SB χ^2, Satorra-Bentler scaled chi-square; df, degrees of freedom; CFI, comparative fit index; RMSEA, root mean square error of approximation.
**$p < 0.001$. Gender ($n_{male} = 236$, $n_{female} = 425$), Age ($n_{under20years-old} = 383$, $n_{20+years-old} = 281$).*

on the general factor. Moreover, 92% of the reliable variance of the total scores could be attributed to the general well-being factor ($\omega_h/\omega = 0.92$) and only 7% of the reliable variance can be attributed to the multidimensionality associated with the specific well-being factors ($\omega - \omega_h = 0.07$). Thus, raw total scores in the Kazakhstani version of the MHC-SF can be interpreted as unidimensional reflections of well-being.

Omega subscale coefficients indicated that 47% of the variance in the EWB subscale could be explained by EWB and the general factor ($\omega_{ewb} = 0.47$), 64% of the variance in the SWB subscale could be explained by SWB and the general factor ($\omega_{ewb} = 0.64$), and 75% of the variance in the PWB subscale could be explained by PWB and the general factor ($\omega_{ewb} = 0.75$). However, the omega hierarchical subscale (ω_{hs}) coefficients were low for the three well-being subscales ($\omega_{h-ewb} = 0.09$, $\omega_{h-swb} = 0.15$, $\omega_{h-pwb} = 0.12$), indicating that the ability of the subscales of the Kazakhstani version of the MHC-SF to reliably measure specific variance of the three well-being components is low.

Measurement Invariance

The MGCFA results for measurement invariance for the Kazakhstani version of the MHC-SF across gender (male vs. female) and age groups (under 20 year-old vs. 20+ year-old) are presented in **Table 3**. The bifactor model fitted the data satisfactorily across gender and age, indicating that configural invariance was supported for both variables (CFI > 0.95, RSMEA < 0.08). Equality constraints were imposed on all factor loadings for both gender and domain groups to test full metric invariance. The ΔCFI and ΔRSMEA indicated full metric invariance (ΔCF < 0.01, ΔRSMEA < 0.30). Equality constraints were then imposed on all intercepts and the three difference tests also indicated full scalar invariance. Finally, equality constraints were imposed on all residual variances, with the ΔCFI and ΔRSMEA supporting full strict invariance. These findings demonstrate the measurement invariance of the Kazakhstani version of the MHC-SF across gender and age.

DISCUSSION AND CONCLUSIONS

The main goal of the present study was to examine the psychometric properties of the Kazakhstani version of the MHC-SF. We used descriptive analysis to examine the distribution of the scores and the prevalence of the categorical diagnoses of states of mental health. Confirmatory factor analysis techniques were implemented to evaluate and compare the fit of the well-being model proposed by the authors of the scale, which comprises three correlated factors representing emotional, social, and psychological well-being (Keyes et al., 2008), with alternative factorial solutions. We also explored the reliability of the scores using hierarchical omega statistics. Moreover, we tested whether the measure varied across different gender and age groups using multigroup confirmatory factor analysis.

Analysis of the distributions of categorical diagnosis of mental health states in our sample demonstrated that 17.6% of the participants reported being languishing, 45.2% moderately mentally healthy, and 27.9% flourishing. These estimates are similar to those found in previous studies with adult samples (Keyes et al., 2008; Lamers et al., 2011; Petrillo et al., 2015) but comparatively lower than those typically reported by younger populations in other contexts (e.g., Luijten et al., 2019). This could be because participants in our study are, on average, older than those in previous studies, and the probabilities of experiencing common psychological challenges increase through adolescence, reaching a peak in early adulthood (Kessler et al., 2007). Also, the data in the present study were collected in the early stage of the COVID-19 pandemic, which obviously contributed to the increased levels of psychological distress in our sample.

The original three-factor model displayed acceptable goodness-of-fit indexes, and comparatively better than those of single-factor and two-factor models. This is consistent with the results reported in previous validation studies of other versions of the MHC-SF (Lamers et al., 2011; Joshanloo et al., 2013; Karaś et al., 2014; Petrillo et al., 2015; Echeverría et al., 2017). However, the results revealed that a bifactor model provides a better representation of the factorial structure of the Kazakhstani version of the MHC-SF compared to the original model and other competing models. These results are consistent with a growing number of studies that suggest that the MHC-SF measures a predominant general well-being factor and three specific factors that correspond to the emotional, social, and psychological well-being subscales (De Bruin and Du Plessis, 2015; Jovanović, 2015; Hides et al., 2016; Echeverría et al., 2017; Rogoza et al., 2018; Longo et al., 2020).

An interesting finding in the present study was that item 4 (i.e., social contribution) had no salient loading on social well-being in the Kazakhstani version of the MHC-SF. Low factor loadings or no statistically significant loadings of item 4 on social well-being have been reported in previous studies using CFA and ESEM approaches in samples from Argentina (Perugini et al., 2017), New Zealand (Joshanloo et al., 2017), Iran (Joshanloo, 2016a), and Serbia (Joshanloo and Jovanović, 2017). Some authors have suggested that this may be because social contribution is normally seen as connected to more individual, private aspects of well-being as it refers to personal "feelings of usefulness" (Bobowik et al., 2015, p. 10). Relatedly, it has been proposed that social contribution may be a more accurate indicator of psychological well-being. In this sense, Joshanloo (2016b) argued that the belief that one can contribute to society is related to the perception that one has a series of facilitating psychological skills, such as positive relationships with others.

The strong loadings of all 14 items on the general well-being factor and the high omega reliability omega hierarchical coefficient (ω_h) confirm the existence of a unitary/cohesive construct of general well-being that is reliably measured by the MHC-SF in the Kazakhstani context. This provides additional support to the proposition that a single overarching well-being construct could accurately integrate several conceptions of well-being (e.g., hedonia and eudaimonia) (Chen et al., 2013; Díaz et al., 2015; Disabato et al., 2016; Strelhow et al., 2020). However, there is relative agreement that the hedonic and eudaimonic components of well-being are related but distinct constructs and correlate differently with other predictors and outcomes (Gallagher et al., 2009; Delle Fave et al., 2011; Huta and Waterman, 2014). Therefore, future research should further explore the adequacy of using a total score to measure the general well-being using other measures and in other contexts.

Despite the good fit of the bifactor model, this study provides limited support for the multidimensionality of the Kazakhstani version of the MHC-SF. First, all items demonstrated statistically significant loadings to their expected specific factors, but these loadings were not substantial and were always lower relative to the general well-being factor. Second, the reliability of the scores was adequate to good for the social and psychological subscales, but low for the emotional subscale, as estimated by omega subscale coefficients (ω_s). Third, the scores for the three well-being subscales explained little variance beyond that explained by the general factor, as indicated by the low omega hierarchical subscale coefficients (ω_{hs}). This indicates that the ability of the Kazakhstani version of the MHC-SF subscales to reliably measure the specific variances of EWB, SWB and PWB is too low, as has been reported for other samples (e.g., Jovanović, 2015).

The findings of the study suggest that the bifactor model could be used to compare parameter estimates across genders. Moreover, an excellent fit was also found when the sample was split into a group of younger and older adults. These conclusions are aligned with previous studies in other contexts and further support the invariance of the structure of the MHC-SF bifactor model across different gender and age groups (e.g., Echeverría et al., 2017; Lamborn et al., 2018).

In sum, the results imply that a bifactor structure of the MHC-SF with a general well-being factor and three specific factors representing EWB, SWB, and PWB fits the data better than the original three factor structure and other competing models in the context of Kazakhstan. Data further supports the adequacy of the bifactor model of the MHC-SF in the sample, regardless of gender or age. The total score of the MHC-SF can be used as a reliable and valid indicator of general well-being that supports the diagnosis of participants as flourishing, moderately mentally healthy, or languishing. However, caution should be applied when using and interpreting the EWB, SWB and PSW subscale scores, as most of the reliable variance in subscale scores is attributable to the general well-being factor. This prevents the interpretation of the specific subscale scores of the Kazakhstani version of the MHC-SF as unequivocal indicators of EWB, SWB, and PSW. A plausible explanation is that the number of items in the MHC-SF could be insufficient to fully capture the complexity and breadth of well-being as a multidimensional construct with two or three correlated subscales—as proposed by the authors of the scale—without accounting for the presence of a general factor of well-being that represents the shared variance among all items (Jovanović, 2015; Longo et al., 2020). Moreover, the number of items in the MHC-SF is perhaps excessive for a single latent factor of well-being to fit adequately the data (Longo et al., 2020). However, it should be noted that bifactor models tend to overfit and set better than any other confirmatory models, regardless of the population's true model (Markon, 2019). Furthermore, bifactor models tend to produce less reliable specific factors that are well represented by their constituent indicators (Watts et al., 2019), as was the case in our study. Therefore, future studies should confirm the suitability of the bifactor structure beyond the interpretation of common fit statistics using innovative approaches, as proposed by Bonifay and Cai (2017).

In this study, we report several approaches to determine the psychometric characteristics of the Kazakhstani version of the MHC-SF. We conclude that this version of the MHC-SF is useful for examining a general factor of well-being and the measurement invariance of the instrument across University gender and age groups. However, we discourage the use of subscale scores as unequivocal indicators of emotional, social, and psychological well-being in this context. Some of the limitations of the present study are the use of self-reported data and the use of convenience University samples. Future research is needed on the dimensionality of Kazakhstani version of the MHC-SF in more representative samples. Another limitation is that the study does not examine the relationships between Kazakhstani version of the MHC-SF scores and convergent measures. Other studies should test the construct validity of the MHC-SF in Kazakhstan and similar contexts.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Nazarbayev University Institutional Research Ethics Committee (IREC). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DH-T and LI contributed to the conception and design of the study, organized the database, performed the statistical analysis, and wrote the first draft of the manuscript. AM, NL, YN, AA, AC, and JS wrote the sections of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Nazarbayev University Faculty-Development Competitive Research Grants Program (Reference Number 240919FD3902).

REFERENCES

Bobowik, M., Basabe, N., and Páez, D. (2015). The bright side of migration: hedonic, psychological, and social well-being in immigrants in Spain. *Soc. Sci. Res.* 51, 1–16. doi: 10.1016/j.ssresearch.2014.09.011

Bonifay, W., and Cai, L. (2017). On the complexity of item response theory models. *Multivariate Behav. Res.* 52, 465–484. doi: 10.1080/00273171.2017.1309262

Brislin, R. W. (1970). Back-translation for cross-cultural research. *J. Cross-Cult. Psychol.* 1, 185–216. doi: 10.1177/135910457000100301

Brown, G. T. (2006). Teachers' conceptions of assessment: validation of an abridged version. *Psychol. Rep.* 99, 166–170. doi: 10.2466/pr0.99.1.166-170

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Modeling* 14, 464–504 doi: 10.1080/10705510701301834

Chen, F. F., Jing, Y., Hayes, A., and Lee, J.M. (2013). Two concepts or two approaches? A bifactor analysis of psychological and subjective well-being. *J. Happiness Stud.* 14, 1033–1068. doi: 10.1007/s10902-012-9367-x

Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Model.* 9, 233–255. doi: 10.1207/S15328007SEM0902_5

Contreras, E. K. P., Castro, S. E. L., Pacheco, G. A. B., Sizer, M. E. A., Keyes, C. L. M., and Medina, W. P. A. (2017). Reliability and validity of the Mental Health Continuum (MHC-SF) in the Ecuadorian contexts. *Ciencias Psicol.* 11, 223–232. doi: 10.22235/cp.v11i2.1499

De Bruin, G. P., and Du Plessis, G. A. (2015). Bifactor analysis of the mental health continuum—Short form (MHC—SF). *Psychol. Rep.* 116, 438–446. doi: 10.2466/03.02.PR0.116k20w6

de Carvalho, J. S., Pereira, N. S., Pinto, A. M., and Marôco, J. (2016). Psychometric properties of the mental health continuum-short form: a study of Portuguese speaking children/youths. *J. Child Fam. Stud.* 25, 2141–2154. doi: 10.1007/s10826-016-0396-7

Delle Fave, A., Brdar, I., Freire, T., Vella-Brodrick, D., and Wissing, M. P. (2011). The eudaimonic and hedonic components of happiness: qualitative and quantitative findings. *Soc. Indic. Res.* 100, 185–207. doi: 10.1007/s11205-010-9632-5

Díaz, D., Stavraki, M., Blanco, A., and Gandarillas, B. (2015). The eudaimonic component of satisfaction with life and psychological well-being in Spanish cultures. *Psicothema* 27, 247–253. doi: 10.7334/psicothema2015.5

Disabato, D. J., Goodman, F. R., Kashdan, T. B., Short, J. L., and Jarden, A. (2016). Different types of well-being? A cross-cultural examination of hedonic and eudaimonic well-being. *Psychol. Assess.* 28:471. doi: 10.1037/pas0000209

Donnelly, A., O'Reilly, A., Dolphin, L., O'Keeffe, L., and Moore, J. (2019). Measurement of the performance of the short-form mental health continuum (MHC-SF) in a primary care youth mental health service. *Ir. J. Psychol. Med.* 36, 201–205. doi: 10.1017/ipm.2018.55

Douglas, S. P., and Craig, C. S. (2007). Collaborative and iterative translation: an alternative approach to back translation. *J. Int. Market.* 15, 30–43. doi: 10.1509/jimk.15.1.030

Echeverría, G., Torres, M., Pedrals, N., Padilla, O., Rigotti, A., and Bitran, M. (2017). Validation of a Spanish version of the mental health continuum-short form questionnaire. *Psicothema* 29, 96–102. doi: 10.7334/psicothema2016.3

Epskamp, S. (2015). semPlot: unified visualizations of structural equation models. *Struct. Equ. Model.* 22, 474–483. doi: 10.1080/10705511.2014.937847

Eshun, S., and Gurung, R. A. (eds.). (2009). *Culture and Mental Health: Sociocultural Influences, Theory, and Practice*. John Wiley & Sons.

Fernando S. (2019). Developing mental health services in the global south. *Int. J. Ment. Health* 48, 338–345. doi: 10.1080/00207411.2019.1706237

Gallagher, M. W., Lopez, S. J., and Preacher, K. J. (2009). The hierarchical structure of well-being. *J. Pers.* 77, 1025–1050. doi: 10.1111/j.1467-6494.2009.00573.x

Guo, C., Tomson, G., Guo, J., Li, X., Keller, C., and Söderqvist, F. (2015). Psychometric evaluation of the Mental Health Continuum-Short Form (MHC-SF) in Chinese adolescents–a methodological study. *Health Qual. Life Outcomes* 13, 1–9. doi: 10.1186/s12955-015-0394-2

Hides, L., Quinn, C., Stoyanov, S., Cockshaw, W., Kavanagh, D. J., Shochet, I., et al. (2020). Testing the interrelationship between mental well-being and mental distress in young people. *J. Positive Psychol.* 15, 314–324. doi: 10.1080/17439760.2019.1610478

Hides, L., Quinn, C., Stoyanov, S., Cockshaw, W., Mitchell, T., and Kavanagh, D. J. (2016). Is the mental wellbeing of young Australians best represented by a single, multidimensional or bifactor model? *Psychiatry Res.* 241, 1–7. doi: 10.1016/j.psychres.2016.04.077

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Huta, V., and Waterman, A. S. (2014). Eudaimonia and its distinction from hedonia: Developing a classification and terminology for understanding conceptual and operational definitions. *J. Happiness Stud.* 15, 1425–1456. doi: 10.1007/s10902-013-9485-0

Jaotombo, F. (2019). Study of the Mental Health Continuum Short Form (MHC-SF) amongst french workers: a combined variable-and person-centered approach. *J. Well-Being Assess.* 3, 97–121. doi: 10.1007/s41543-019-00022-z

Joshanloo, M. (2016a). A new look at the factor structure of the MHC-SF in Iran and the United States using exploratory structural equation modeling. *J. Clin. Psychol.* 72, 701–713. doi: 10.1002/jclp.22287

Joshanloo, M. (2016b). Revisiting the empirical distinction between hedonic and eudaimonic aspects of well-being using exploratory structural equation modeling. *J. Happiness Stud.* 17, 2023–2036. doi: 10.1007/s10902-015-9683-z

Joshanloo, M. (2018). The structure of the MHC-SF in a large American sample: contributions of multidimensional scaling. *J. Ment. Health* 26, 139–146. doi: 10.1080/09638237.2018.1466044

Joshanloo, M. (2019). Factor structure and measurement invariance of the MHC-SF in the USA. *Eur. J. Psychol. Assess.* 35, 521–525. doi: 10.1027/1015-5759/a000425

Joshanloo, M. (2020). Factorial/Discriminant validity and longitudinal measurement invariance of MHC-SF in Korean young adults. *Curr. Psychol.* 39, 51–57. doi: 10.1007/s12144-017-9742-1

Joshanloo, M., Jose, P. E., and Kielpikowski, M. (2017). The value of exploratory structural equation modeling in identifying factor overlap in the Mental Health Continuum-Short Form (MHC-SF): a study with a New Zealand sample. *J. Happiness Stud.* 18, 1061–1074. doi: 10.1007/s10902-016-9767-4

Joshanloo, M., and Jovanović, V. (2017). The factor structure of the mental health continuum-short form (MHC-SF) in Serbia: an evaluation using exploratory structural equation modeling. *J. Ment. Health* 26, 510–515. doi: 10.1080/09638237.2016.1222058

Joshanloo, M., and Lamers, S. M. (2016). Reinvestigation of the factor structure of the MHC-SF in the Netherlands: contributions of exploratory structural equation modeling. *Pers. Individ. Differ.* 97, 8–12. doi: 10.1016/j.paid.2016.02.089

Joshanloo, M., Wissing, M. P., Khumalo, I. P., and Lamers, S. M. (2013). Measurement invariance of the Mental Health Continuum-Short Form (MHC-SF) across three cultural groups. *Pers. Indiv. Differ*. 55, 755–759. doi: 10.1016/j.paid.2013.06.002

Jovanović, V. (2015). Structural validity of the Mental Health Continuum-Short Form: The bifactor model of emotional, social and psychological well-being. *Pers. Individ. Differ*. 75, 154–159. doi: 10.1016/j.paid.2014.11.026

Karaś, D., Cieciuch, J., and Keyes, C. L. (2014). The polish adaptation of the mental health continuum-short form (MHC-SF). *Pers. Individ. Differ*. 69, 104–109. doi: 10.1016/j.paid.2014.05.011

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., and Ustun, T. B. (2007). Age of onset of mental disorders: a review of recent literature. *Curr. Opin. Psychiatr.* 20, 359–364. doi: 10.1097/YCO.0b013e32816ebc8c

Keyes, C. L. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *J. Consult. Clin. Psychol*. 73, 539–548. doi: 10.1037/0022-006X.73.3.539

Keyes, C. L. (2007). Promoting and protecting mental health as flourishing: a complementary strategy for improving national mental health. *Am. Psychol.* 62, 95–108. doi: 10.1037/0003-066X.62.2.95

Keyes, C. L., Wissing, M., Potgieter, J. P., Temane, M., Kruger, A., and Van Rooy, S. (2008). Evaluation of the mental health continuum–short form (MHC–SF) in setswana-speaking South Africans. *Clin. Psychol. Psychother*. 15, 181–192. doi: 10.1002/cpp.572

Keyes, C. L. M. (2002). The Mental Health Continuum: from languishing to flourishing in life. *J. Health Soc. Behav*. 43, 207–222. doi: 10.2307/3090197

Keyes, C. L. M. (2006). Mental health in adolescence: is America's youth flourishing? *AJO* 76, 395–402. doi: 10.1037/0002-9432.76.3.395

Keyes, C. L. M. (2013). *Mental Well-Being: International Contributions to the Study of Positive Mental Health.* Dordrecht, The Netherlands: Springer. doi: 10.1007/978-94-007-5195-8

Kobau, R., Seligman, M. E., Peterson, C., Diener, E., Zack, M. M., Chapman, D., et al. (2011). Mental health promotion in public health: perspectives and strategies from positive psychology. *AJPH* 101, e1–e9. doi: 10. 2105/AJPH.2010.300083

Lamborn, P., Cramer, K. M., and Riberdy, A. (2018). The structural validity and measurement invariance of the mental health continuum–short form (MHC-SF) in a large Canadian sample. *J Well-Being Assess*. 2, 1–19. doi: 10.1007/s41543-018-0007-z

Lamers, S. M., Westerhof, G. J., Bohlmeijer, E. T., ten Klooster, P. M., and Keyes, C. L. (2011). Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF). *J. Clin. Psychol*. 67, 99–110. doi: 10.1002/jclp.20741

Lim, Y. J. (2014). Psychometric characteristics of the Korean Mental Health Continuum–Short Form in an adolescent sample. *J. Psychoeduc. Assess*. 32, 356–364. doi: 10.1177/0734282913511431

Longo, Y., Jovanović, V., Sampaio de Carvalho, J., and Karaś, D. (2020). The general factor of well-being: multinational evidence using bifactor ESEM on the Mental Health Continuum–Short Form. *Assessment* 27, 596–606. doi: 10.1177/1073191117748394

Luijten, C. C., Kuppens, S., van de Bongardt, D., and Nieboer, A. P. (2019). Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF) in Dutch adolescents. *Health Qual. Life Outcomes* 17:157. doi: 10.1186/s12955-019-1221-y

Markon, K. E. (2019). Bifactor and hierarchical models: Specification, inference, and interpretation. *Annu. Rev. Clin. Psychol*. 15, 51–69. doi: 10.1146/annurev-clinpsy-050718-095522

Monteiro, F., Fonseca, A., Pereira, M., and Canavarro, M. C. (2021). Measuring positive mental health in the postpartum period: the bifactor structure of the mental health Continuum–Short Form in Portuguese Women. *Assessment* 28, 1434–1444. doi: 10.1177/1073191120910247

Perreira, T. A., Morin, A. J. S., Hebert, M., Gillet, N., Houle, S. A., and Berta, W. (2018). The short form of the Workplace Affective Commitment Multidimensional Questionnaire (WACMQ-S): a bifactor-ESEM approach among healthcare professionals. *J. Vocat. Behav.* 106, 62–83. doi: 10.1016/j.jvb.2017.12.004

Perugini, M. L. L., de la Iglesia, G., Solano, A. C., and Keyes, C. L. M. (2017). The mental health continuum–short form (MHC–SF) in the Argentinean context: confirmatory factor analysis and measurement invariance. *Eur. J. Psychol*. 13:93. doi: 10.5964/ejop.v13i1.1163

Petrillo, G., Capone, V., Caso, D., and Keyes, C. L. (2015). The mental health continuum–short form (MHC–SF) as a measure of well-being in the Italian context. *Soc. Indic. Res.* 121, 291–312. doi: 10.1007/s11205-014-0629-3

Putnick, D. L., and Bornstein, M. H. (2016). Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev Rev*. 41, 71–90. doi: 10.1016/j.dr.2016.06.004

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/ (accessed June 8, 2021).

Rafiey, H., Alipour, F., LeBeau, R., Amini Rarani, M., Salimi, Y., and Ahmadi, S. (2017). Evaluating the psychometric properties of the Mental Health Continuum-Short Form (MHC-SF) in Iranian earthquake survivors. *IJMHS* 46, 243–251. doi,: 10.1080/00207411.2017.1308295

Revelle, W (2021). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.1.3. Available online at: https://CRAN.R-project.org/package=psych (accessed June 8, 2021).

Rodriguez, A., Reise, S. P., and Haviland, M. G. (2016). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychol. Methods* 21, 137–150. doi: 10.1037/met0000045

Rogoza, R., Truong Thi, K. H., Rózycka-Tran, J., Piotrowski, J., and Zemojtel-Piotrowska, M. (2018). Psychometric properties of the MHC-SF: an integration of the existing measurement approaches. *J. Clin. Psychol*. 74, 1742–1758. doi: 10.1002/jclp.22626

Rosseel, Y. (2012). Lavaan: an R package for structural equation modeling. *J. Stat. Softw*. 48, 1–36. doi: 10.18637/jss.v048.i02

Ryan, R. M., and Deci, E. L. (2001). On happiness and human potentials: a review of research on hedonic and eudaemonic well-being. *Annu. Rev. Psychol*. 52, 141–166. doi: 10.1146/annurev.psych.52.1.141

Santini, Z. I., Jose, P. E., Koyanagi, A., Meilstrup, C., Nielsen, L., Madsen, K. R., et al. (2020). Formal social participation protects physical health through enhanced mental health: a longitudinal mediation analysis using three consecutive waves of the Survey of Health, Ageing and Retirement in Europe (SHARE). *Soc. Sci. Med*. 251:112906. doi: 10.1016/j.socscimed.2020.112906

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., and King, J. (2006). Reporting structural equation modelling and confirmatory factor analysis results: a review. *J. Educ. Res*. 99, 323–338. doi: 10.3200/JOER.99.6.323-338

Schutte, L., and Wissing, M. P. (2017). Clarifying the factor structure of the mental health continuum short form in three languages: a bifactor exploratory structural equation modeling approach. *Soc. Ment. Health* 7, 142–158. doi: 10.1177/2156869317707793

Silverman, A. L., Forgeard, M., Beard, C., and Björgvinsson, T. (2018). Psychometric properties of the mental health continuum–short form in a psychiatric sample. *J. Well-Being Assess*. 2, 57–73. doi: 10.1007/s41543-018-0011-3

Strelhow, M. R. W., Castellá Sarriera, J. C., and Casas, F. (2020). Evaluation of well-being in adolescence: proposal of an integrative model with hedonic and eudemonic aspects. *Child Indic. Res*. 13, 1439–1452. doi: 10.1007/s12187-019-09708-5

Vaillant, G. E. (2012). Positive mental health: is there a cross-cultural definition?. *World Psychiatry* 11, 93–99. doi: 10.1016/j.wpsyc.2012.05.006

van Erp Taalman Kip, R. M., and Hutschemaekers, G. J. (2018). Health, well-being, and psychopathology in a clinical population: structure and discriminant validity of Mental Health Continuum Short Form (MHC-SF). *J. Clin. Psychol.* 74, 1719–1729. doi: 10.1002/jclp.22621

Watts, A. L., Poore, H. E., and Waldman, I. D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clin. Psychol. Sci*. 7, 1285–1303. doi: 10.1177/2167702619855035

World Health Organization (2004). *Promoting Mental Health: Concepts, Emerging Evidence, Practice (Summary Report)*. Geneva: World Health Organization. Available online at: https://www.who.int/mental_health/evidence/en/promoting_mhh.pdf (accessed April 28, 2021).

Zemojtel-Piotrowska, M., Piotrowski, J. P., Osin, E. N., Cieciuch, J., Adams, B. G., Ardi, R., et al. (2018). The mental health continuum-short form: the structure and application for cross-cultural studies—a 38 nation study. *J. Clin. Psychol.* 74, 1034–1052. doi: 10.1002/jclp.22570

Zinbarg, R. E., Revelle, W., Yovel, I., and Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω H: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika* 70, 123–133. doi: 10.1007/s11336-003-0974-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 November 2021
doi: 10.3389/fpsyg.2021.783761

Check for updates

The Swedish Version of the Multidimensional Inventory for Religious/Spiritual Well-Being: First Results From Swedish Students

*Magdalena Wenzl[1], Jürgen Fuchshuber[2,3], Nikita Podolin-Danner[1], Giorgia Silani[1] and Human-Friedrich Unterrainer[2,4,5]**

[1]Faculty of Psychology, University of Vienna, Vienna, Austria, [2]CIAR: Center for Integrative Addiction Research, Grüner Kreis Society, Vienna, Austria, [3]Department of Philosophy, University of Vienna, Vienna, Austria, [4]Department of Psychiatry and Psychotherapeutic Medicine, Medical University of Graz, Graz, Austria, [5]Department of Religious Studies, University of Vienna, Vienna, Austria

Background: Studies investigating the relationship between religiosity/spirituality and mental health have suggested both positive and negative associations, highlighting the importance of multifaceted assessment of these rather broad constructs. The present study aims at contributing to this field of research by providing a validated Swedish version of the Multidimensional Inventory for Religious/Spiritual Well-Being (MI-RSWB-S) and further examining how this instrument relates to Big Five personality factors, Sense of Coherence (SOC), and religiosity.

Methods: Data were collected from a total of 1,011 Swedish students (747 females; age range 18–40) *via* completion of an online survey, including a new Swedish Version of the MI-RSWB-S, the Ten Item Personality Inventory (TIPI), the Sense of Coherence Scale (SOC-13), and the Centrality of Religiosity Scale (CRS-5).

Results: Results revealed adequate estimates of internal consistency and substantial evidence for the postulated six-dimensional structure. However, confirmatory factor analysis yielded poor fit indices, resulting in the development and validation of a revised measure of Religious/Spiritual Well-Being (RSWB), comprising the subscales General Religiosity and Connectedness. Most of the MI-RSWB-S dimensions were positively correlated with the personality domains Extraversion, Openness to Experience, Conscientiousness, and Agreeableness and negatively related to Neuroticism. SOC was positively linked to Hope Immanent, Forgiveness, Hope Transcendent, and Experiences of Sense of Meaning, whereas CRS exhibited positive correlations with all MI-RSWB-S subscales except Hope Transcendent.

Conclusion: The findings of the current study support the validity and reliability of the Swedish adoption of the MI-RSWB and confirm previously reported associations with the Big Five personality traits, SOC, and CRS. More in general, our results underline the putative substantial link between RSWB dimensions and mental health. Further research especially in clinical surroundings as well as by employing more representative samples is now warranted.

Keywords: religiosity, spirituality, personality factors, sense of coherence, scale development

OPEN ACCESS

Edited by:
Paola Gremigni,
University of Bologna, Italy

Reviewed by:
Mahboubeh Dadfar,
Iran University of Medical Sciences, Iran
Helmar Bornemann-Cimenti,
Medical University of Graz, Austria

Correspondence:
Human-Friedrich Unterrainer
human.unterrainer@univie.ac.at

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *26 September 2021*
Accepted: *19 October 2021*
Published: *09 November 2021*

Citation:
Wenzl M, Fuchshuber J, Podolin-Danner N, Silani G and Unterrainer H-F (2021) The Swedish Version of the Multidimensional Inventory for Religious/Spiritual Well-Being: First Results From Swedish Students. Front. Psychol. 12:783761. doi: 10.3389/fpsyg.2021.783761

INTRODUCTION

Sweden has been described as one of the most secular countries in the world (Esmer and Pettersson, 2007), with low levels of church attendance and belief in God, heaven and hell, the existence of sin, salvation, life after death, and the divinity of the Bible (Zuckerman, 2020). In a study comparing the importance of religion, worship attendance, frequency of prayer, and belief in God in 34 European countries, Sweden was ranked among the bottom 10 countries in all categories. Only 10% of the Swedish respondents stated that religion is very important in their lives, and about the same percentage of subjects claimed to pray daily (11%) and attend worship services at least monthly (11%). About 14% expressed that they believe in God with absolute certainty. Using an overall combined index, Sweden came in 30th place, with 10% of the sample defined as highly religious (Pew Research Center, 2018a). Consistent with global observations, Swedish men and young adults tend to be less religious than their female and older counterparts (Pew Research Center, 2016, 2018b). Despite low levels of religiosity and religious participation, accompanied by a continuous decline in membership, the Church of Sweden has approximately 5.7 million members, accounting for more than half of the Swedish population (Kyrkan, 2021). Moreover, although Swedish people generally hold sceptical and critical attitudes towards religion, relatively few call themselves convinced atheists. Many believe in "something" and possess an individualistic, non-dogmatic, and spiritual outlook on life (Thurfjell, 2019). The majority of the ones who do believe in God refer to a vague, distant entity rather than to biblical conceptions of God as an absolute, vengeful, or merciful being (Zuckerman, 2020).

Studies examining associations between religiosity/spirituality (R/S) and psychological health have yielded mixed results, linking R/S to various parameters of increased (e.g., high levels of well-being and meaning and purpose) as well as decreased mental health or mood pathology (e.g., cognitive rigidity, excessive concern over sins, and delayed professional psychological treatment; Koenig et al., 2012). The relationship between R/S and mental health is substantially dependent on the operationalisation of these rather "fuzzy" constructs (Zinnbauer et al., 1997; see also Hodapp and Zwingmann, 2019). Hence, its thorough investigation calls for the utilisation of instruments designed to capture the complexity of R/S contents (Koenig, 2008).

The Multidimensional Inventory for Religious/Spiritual Well-Being (MI-RSWB) is a multidimensional instrument for the assessment of Religious/Spiritual Well-Being (RSWB), defined as "the ability to experience and integrate meaning and purpose in existence through a connectedness with self, others, or a power greater than oneself" (Unterrainer et al., 2011, p. 117). The MI-RSWB comprises 48 items equally divided into six subscales: General Religiosity (GR), Connectedness (CO), Hope Transcendent (HT), Hope Immanent (HI), Forgiveness (FO), and Experiences of Sense and Meaning (SM). The dimension GR relates to traditional religious beliefs and institutionalised religion. CO refers to spiritual inclination and the feeling of being immersed in something bigger than oneself. The subscale HT reflects one's hope for a better afterlife, whereas HI describes the extent to which one is hopeful for a better future here on earth. FO refers to the ability to extend forgiveness to oneself and others and to resign oneself to things that have gone wrong. The dimension SM pertains to meaningful life experiences, including those of honesty, gratitude, and true friendship. The first three subscales (GR, CO, and HT) can be used as parameters for the transcendent area of well-being [sub-score Transcendent Well-Being (TWB)], while the remaining subscales (HI, FO, and SM) relate to the immanent area of perception [sub-score Immanent Well-Being (IWB)]. In addition, the six dimensions can be summarised into a total score, thus providing a global measure of RSWB (Unterrainer et al., 2012).

The development of the MI-RSWB was initiated as a response to R/S needs of clinical and non-clinical populations and involved theoretical considerations in terms of an integration of a R/S dimension in the biopsychosocial model of health and illness (Engel, 1977). Furthermore, the MI-RSWB may be thought of as a multidimensional alternative to the two-dimensional Spiritual Well-Being Scale (SWBS), originated by Ellison (1983) (see also Moberg, 1971). The subscales of the SWBS, Religious Well-Being (RWB), and Existential Well-Being (EWB), are reflected in the differentiation of TWB and IWB. However, the inclusion of six dimensions allows for a more detailed examination of these areas of perception. The very content of these subscales was a result of interdisciplinary group discussion, literature research, and statistical analysis of empirical data (Unterrainer, 2021).

The original Austrian–German version of the inventory has been applied in various clinical and non-clinical studies. The latter resulted in norm values for the Austrian general population, with appealing psychometric properties for the subscales and the total score (Unterrainer and Fink, 2013). So far, the questionnaire has been translated into and validated in English (Unterrainer et al., 2012), Italian (Stefa-Missagli et al., 2014), Mexican Spanish (Berger et al., 2016), Bosnian (Malinovic et al., 2016), Russian (Agarkov et al., 2018), and Farsi (Dadfar et al., 2019), thereby demonstrating satisfactory psychometric properties (with one exception for HT in Dadfar et al., 2019). However, confirmatory factor analysis of data obtained in Austrian (Unterrainer et al., 2010) and British samples (Unterrainer et al., 2012) revealed only limited empirical support for the original MI-RSWB structure.

The MI-RSWB has been related to a number of psychosocial measures, including the "Big Five"-Factor Model (FFM), as most prominently described by Costa and McCrae (1992a). Studies investigating its associations with these personality domains have found positive correlations between extraversion and RSWB, HI, and SM (e.g., Berger et al., 2016) and negative links between Neuroticism and RSWB, whereas GR, CO, and SM proved to be unrelated to this specific trait (e.g., Malinovic et al., 2016). The personality dimension Openness to Experience has been consistently positively linked with SM and, in some cases, with the MI-RSWB total score, GR, HI, and CO (e.g., Hiebler-Ragger et al., 2018). Moreover, positive relationships have been noted between all MI-RSWB measures and both conscientiousness and agreeableness (e.g., Unterrainer et al., 2012). Similar results have been reported in other studies examining the relationship between R/S and these two personality traits (see Saroglou, 2010). What is more, substantial positive correlations have been observed between sense of coherence (SOC) and all MI-RSWB dimensions but CO, with the strongest association found for HI (e.g., Unterrainer et al., 2010). Finally, the MI-RSWB has been proven to be significantly related to other prominent measures of religiosity, such as the "Centrality of Religiosity" C-Scale (CRS; Huber and Huber, 2012). Thereby, for instance, Berger et al. (2016) reported positive links between CRS and all MI-RSWB scores except HT.

The current study aims are two-fold: As a first step, it is intended to introduce an internally validated Swedish version of the MI-RSWB-S. As a second step (external validation), it is planned to relate the MI-RSWB-S to established measures of the Big Five personality traits, SOC, and religiosity.

MATERIALS AND METHODS

Participants and Procedure

This study is based on a convenience sample of Swedish students. The inclusion criteria for the study were: (1) Swedish citizenship, (2) fluency in the Swedish language, (3) student enrolment at a Swedish university or university college, and (4) being between 18 and 40 years of age. Furthermore, a minimum survey completion time criterion of 4 min was implemented, leading to the exclusion of one respondent.

Data were acquired between March 8 and April 5, 2021, by means of an online survey, using the web-based software tool SoSci Survey. Participants were primarily recruited through Facebook groups (> 230 groups) and Instagram accounts connected to Swedish universities and university colleges. These were instructed to download a PDF file containing comprehensive information about the study and indicated their consent to participate in it by checking a "yes" box before gaining access to the survey. Subjects were not compensated for their involvement in the research project. Ethical approval for the study was granted by the Ethics Committee of the University of Vienna.

Psychometric Assessment

Multidimensional Inventory for Religious/Spiritual Well-Being

The original Austrian–German version of the MI-RSWB (Unterrainer et al., 2010) was translated into Swedish by a native Swedish-speaking psychology student fluent in German (M.W.). A back-translation was provided by a Swedish–German bilingual speaker, showing a high level of equivalence with the original instrument. The Swedish adoption of the MI-RSWB (MI-RSWB-S) comprises 48 items rated using a six-point Likert scale ranging from 1 (totally disagree) to 6 (totally agree). The items are equally distributed among the six subscales GR (e.g., "My faith gives me a feeling of security."), FO (e.g., "There are things which I cannot forgive," with reverse coding), HI (e.g., "I view the future with optimism."), CO (e.g., "I have experienced the feeling of being absorbed into something greater."), HT (e.g., "I often think about the fact that I will have to leave behind my loved ones.," with reverse coding), and Experiences of SM (e.g., "I have experienced true (authentic) feelings."). The total list of items of the MI-RSWB-S is presented in **Supplementary Material**.

High internal consistency has been reported for the original scale ($\alpha = 0.89$ for the total RSWB score and $\alpha \geq 0.73$ for the subscales; Unterrainer et al., 2010) as well as for the English (Unterrainer et al., 2012), Italian (Stefa-Missagli et al., 2014), Mexican Spanish (Berger et al., 2016), Bosnian (Malinovic et al., 2016), and Russian adoptions (Agarkov et al., 2018), revealing Cronbach's α coefficients of at least 0.83 with respect to the total score.

Ten Item Personality Inventory

The Ten Item Personality Inventory (TIPI) is a brief measure of the Big Five personality domains of the Five-Factor Model (FFM; Costa and McCrae, 1992a), namely Extraversion, Neuroticism, Openness to Experience, Conscientiousness, and Agreeableness. Extraversion includes traits related to sociability, activity, and positive affectivity, while Neuroticism describes the individual tendency to experience psychological distress. Openness to Experience refers to the extent to which an individual is intellectually curious, behaviourally flexible, emotionally differentiated, and non-dogmatic as well as sensitive to imagination, art, and beauty. Conscientiousness reflects the degree to which a person is scrupulous, well-organized, and diligent. Individuals who score high on Agreeableness are trusting, sympathetic, and cooperative, whereas people with a low level of Agreeableness tend to be cynical, callous, and antagonistic.

The TIPI assesses Extraversion, Emotional Stability (reversed Neuroticism), Openness to Experience, Conscientiousness, and Agreeableness with two items per personality dimension (e.g., "I see myself as extraverted, enthusiastic" for Extraversion). All items are responded to on a Likert scale ranging from 1 (disagree strongly) to 7 (agree strongly). Owing to the small number of items, some of the TIPI subscales have demonstrated low internal consistency. The highest Cronbach's α coefficient has been obtained for Emotional Stability ($\alpha = 0.73$), followed by Extraversion ($\alpha = 0.68$), Conscientiousness ($\alpha = 0.50$), Openness to Experience ($\alpha = 0.45$), and Agreeableness ($\alpha = 0.40$; Gosling et al., 2003). However, the TIPI has reached satisfactory levels of convergent and discriminant validity when related to the Big-Five Inventory (BFI; John and Srivastava, 1999), thereby displaying convergences ($r = 0.65$–0.87; $p < 0.01$; mean $r = 0.77$) comparable to those of the well-established multi-item instruments Trait Descriptive Adjectives (TDA; Goldberg, 1992; mean $r = 0.81$) and NEO Five-Factor Inventory (NEO-FFI; Costa and McCrae, 1992b; mean $r = 0.73$), and discriminant correlations below 0.37 (absolute mean $r = 0.20$). Furthermore, the subscales have shown adequate test–retest reliability ($r = 0.62$–0.77; mean $r = 0.72$; Gosling et al., 2003). The Swedish version of the TIPI has been provided by Lundell (2014).

Sense of Coherence Scale

The Sense of Coherence 13-item scale (SOC-13) is a short version of the original Sense of Coherence Scale (SOC-29) used to measure levels of Sense of Coherence (SOC), the core concept of Antonovsky's salutogenic model (Antonovsky, 1987). SOC pertains to a global orientation which expresses an individual's ability to cope with stress and to stay healthy (Mittelmark et al., 2017). It comprises three components: Comprehensibility, Manageability, and Meaningfulness. Comprehensibility refers to the perception that internal and external stimuli are structured, predictable, and explicable, while Manageability reflects the perceived availability of resources to deal with the demands presented by the stimuli. Meaningfulness describes the extent to which these demands are seen as challenges worthy of personal commitment. In a broader sense, Meaningfulness relates to the feeling that life makes sense and has emotional meaning (Antonovsky, 1987, 1991b).

The SOC-13 utilises a seven-point Likert scale with varying verbal response anchors to capture Comprehensibility (five items), Manageability (four items), and Meaningfulness (four items; e.g., "Do you have the feeling that you really do not care about what is going on around you?"). The internal consistency of the SOC-13 scale has been investigated in an exhaustive amount of studies, with Cronbach's α values ranging from 0.70 to 0.92 (Eriksson and Lindström, 2005). Swedish studies have reported good internal consistency for the general population (e.g., Larsson and Kallenberg, 1999) as well as for adolescents and young adults (Räty et al., 2003). The Swedish translation of the SOC-13 (KASAM-13) has been published in Antonovsky (1991a).

Centrality of Religiosity Scale

The Centrality of Religiosity Scale (CRS) is an instrument for assessing the centrality, importance, or salience of religious meanings in personality (Huber and Huber, 2012). It captures five core dimensions of religiosity, namely Public Practice, Private Practice, Religious Experience, Ideology, and Intellect. Public Practice reflects the extent to which one integrates their religious life in a social organism by for instance participating in public religious rituals or activities, whereas Private Practice relates to activities and rituals of personal devotion to a transcendent sphere of reality (e.g., prayer and meditation). The dimension of religious experience includes individual experiences and feelings of being connected to an ultimate reality. Ideology refers to religious beliefs, convictions, and patterns of plausibility (e.g., with respect to the existence of God), while Intellect pertains to religious knowledge and interest, hermeneutical skills, and ways of thinking.

The Swedish version of the CRS-5 (CRS-5 SWE) has been provided by Sjöborg (2014). It measures the five dimensions of religiosity with one item each. These items are rated on a five-point (1–5; Ideology, Intellect, and Religious Experience), six-point (1–6; Public Practice), or eight-point Likert scale (1–8; Private Practice; "How often do you pray?") with different verbal response anchors, together with a "do not know" option. Data collected by means of six-point and eight-point response formats are recoded into values between 1 and 5. The composite score of the CRS-5 has demonstrated high internal consistency ($\alpha = 0.85$; Huber and Huber, 2012).

Statistical Analysis

Data were analysed in three stages. First, a principal component analysis (PCA) with VARIMAX rotation was conducted using the first 300 responses (exploration phase sample). As a second step, confirmatory factor analyses (CFA) were carried out on a different sample, as suggested by Boateng et al. (2018), comprising the remaining data set ($n = 711$; validation phase sample). Model fit was considered acceptable if the following criteria were met: (a) Comparative Fit Index (CFA) > 0.90, (b) Tucker–Lewis Index (TLI) > 0.90, (c) Normed Fit Index (NFI) > 0.90, and (d) Square Root Error Approximation (RMSEA) < 0.08 with the upper bound of the 90% CI < 0.10 (Kline, 2016). Third, descriptive statistics of the total sample ($N = 1,011$) were generated to provide an overview of the MI-RSWB-S scores. In addition, independent *t* tests and Pearson's correlations were performed to examine gender and age effects and how the MI-RSWB-S measures relate to each other and the validation scales (TIPI, SOC-13, and CRS-5). Cronbach's α coefficients were calculated to determine the internal consistency of these instruments, following the guidelines provided by George and Mallery (2016). PCA, descriptive statistics, *t* tests, reliability analysis, and Pearson's correlations were conducted *via* SPSS 25, whereas the CFAs were computed using AMOS 26.

RESULTS

Sample Characteristics

A total of 1,011 participants were included in the final study sample. The sociodemographic characteristics of the total sample ($N = 1,011$), exploration phase sample ($n = 300$), and the validation phase sample ($n = 711$) are given in **Table 1**. Almost three-quarters (73.9%) of the subjects in the total sample identified as female, and more than three-quarters (77.9%) were between the ages of 20 and 29 ($M = 24.78$; $SD = 4.83$). Nine hundred and two respondents (89.2%) were born in Sweden. All 21 counties were represented in the sample, with a relatively high proportion of subjects coming from Stockholm (21.6%) and Västra Götaland (20.6%). Data were obtained from students from a total of 36 university/university colleges (e.g., the University of Gothenburg, Uppsala University, and Stockholm University). Most of the participants were single (45.9%), living together with their partner (22.9%) or in a relationship (18.2%). Less than 10 percent (9.5%) had biological children ($M = 0.19$; $SD = 0.65$). Almost half of the respondents (46.2%) were members of the Church of Sweden, and approximately one-sixth (16.6%) belonged to another Christian denomination. About the same number of subjects had never been part of (16.4%) or left (15.4%) a religious community. Less than 10 percent (7.0%) labelled themselves as Muslims, predominately as adherents to Sunni

TABLE 1 | Sample characteristics.

Variable	Total sample ($N = 1{,}011$) *n* (%)	Exploration phase sample ($n = 300$) *n* (%)	Validation phase sample ($n = 711$) *n* (%)
Gender			
Female	747 (73.9)	217 (72.3)	530 (74.5)
Male	252 (24.9)	79 (26.3)	173 (24.3)
Other	12 (1.2)	4 (1.3)	8 (1.1)
Age			
18–19	63 (6.2)	22 (7.3)	41 (5.8)
20–24	525 (51.9)	142 (47.3)	383 (53.9)
25–29	263 (26.0)	81 (27.0)	182 (25.6)
30–34	100 (9.9)	36 (12.0)	64 (9.0)
35–40	60 (5.9)	19 (6.3)	41 (5.8)
Place of birth			
East Sweden (SE1)	355 (35.1)	93 (31.0)	262 (36.8)
South Sweden (SE2)	391 (38.7)	116 (38.7)	275 (38.7)
North Sweden (SE3)	156 (15.4)	66 (22.0)	90 (12.7)
Other Nordic country	11 (1.1)	2 (0.7)	9 (1.3)
Other European country (outside the Nordics)	35 (3.5)	7 (2.3)	28 (3.9)
Other country (outside Europe)	63 (6.2)	16 (5.3)	47 (6.6)
Relationship status			
Married	91 (9.0)	26 (8.7)	65 (9.1)
Engaged	45 (4.5)	13 (4.3)	32 (4.5)
Cohabitation	232 (22.9)	74 (24.7)	158 (22.2)
In a relationship	184 (18.2)	59 (19.7)	125 (17.6)
Single	464 (45.9)	128 (42.7)	336 (47.3)
Divorced	8 (0.8)	4 (1.3)	4 (0.6)
Religious affiliation			
Christianity (Church of Sweden)	467 (46.2)	145 (48.3)	322 (45.3)
Christianity (free church)	102 (10.0)	36 (12.0)	66 (9.3)
Christianity (Eastern Orthodox)	33 (3.3)	5 (1.7)	28 (3.9)
Christianity (Roman Catholic)	30 (3.0)	5 (1.7)	25 (3.5)
Christianity (other)	3 (0.3)	0 (0.0)	3 (0.4)
Islam (Sunni)	59 (5.8)	10 (3.3)	49 (6.9)
Islam (Shia)	10 (1.0)	2 (0.7)	8 (1.1)
Islam (other)	2 (0.2)	1 (0.3)	1 (0.1)
Judaism	13 (1.3)	6 (2.0)	7 (1.0)
Buddhism	13 (1.3)	3 (1.0)	10 (1.4)
Hinduism	2 (0.2)	0 (0.0)	2 (0.3)
Never been a part of a religious community	166 (16.4)	55 (18.3)	111 (15.6)
Left a religious community	156 (15.4)	50 (16.7)	106 (14.9)
Other	10 (1.0)	1 (0.3)	9 (1.3)

SE1, SE2, and SE3 refer to the first-level nomenclature of territorial units for statistics (NUTS) regions of Sweden.

Islam. Relatively few belonged to a Jewish (1.3%), Buddhist (1.3%), or Hindu (0.2%) community. The participants using the category “others” (1.0%) described their religious affiliation as follows: Druze, Forn Sed, Luciferian, Satanic Temple, Mandaeism, Sikhism, Yogi, and Wicca.

Principal Component Analysis

A PCA with orthogonal rotation (VARIMAX with Kaiser normalisation) was conducted on the 48 items of the MI-RSWB-S in the exploration phase sample ($n = 300$). Based on theoretical considerations regarding the dimensional structure of the MI-RSWB, the number of extracted components was set to 6. Sampling adequacy was evaluated through assessment of the Kaiser-Meyer-Olkin (KMO) index and Bartlett’s test of sphericity. The KMO measure was 0.89, well above the recommended threshold of 0.60 (Kaiser, 1974), and all KMO values for individual items were greater than 0.62, thus exceeding the acceptable limit of 0.50 (Kaiser and Rice, 1974). Furthermore, Bartlett’s test of sphericity was significant ($\chi^2_{(1128)} = 8065.11$, $p < 0.001$), indicating suitability of the data for PCA. As given in **Table 2**, the six-component solution accounted for 53.14% of the total variance. GR explained the largest proportion of the variance.

Confirmatory Factor Analysis

As shown in **Figure 1**, a CFA was performed on the validation phase sample ($n = 711$) to test the factorial structure of the MI-RSWB-S. The CFA yielded incremental fit indices below the acceptable level of 0.90 ($CFA = 0.82$; $TLI = 0.70$; $NFI = 0.81$) and a RMSEA value exceeding 0.08 with a 90% CI upper limit greater than 0.10 ($RMSEA = 0.15$; 90% CI = 0.13–0.16), indicating poor fit. Moreover, RSWB [hereinafter referred to as RSWB Original (RSWB-O)] was unrelated to HT ($\beta = -0.02$) and weakly linked to HI ($\beta = 0.28$) and FO ($\beta = 0.30$; $p < 0.001$ for all values). These results provide little support for the proposed model, prompting further analysis.

Based on conceptual considerations and empirical findings, we developed a revised model (see **Figure 2**) in which GR and CO are summarised into RSWB [RSWB Revised (RSWB-R)], while HI, SM, FO, and HT operate as independent factors. The CFA for the revised model resulted in incremental fit indices above 0.90 ($CFI = 0.995$; $TLI = 0.980$; $NFI = 0.990$) and a RMSEA below 0.08, with a 90% CI upper bound smaller than 0.10 ($RMSEA = 0.039$; 90% CI = 0.000–0.077). In addition, the Chi-square test was non-significant ($\chi 2 = 8.42$; $p > 0.05$; $\chi 2/df = 17.71$). Furthermore, RSWB-R displayed strong associations with GR ($\beta = 0.77$) and CO ($\beta = 0.83$; $p < 0.001$ for all calculations). Taken together, these findings lend support for the postulated structure.

Descriptive Statistics

Table 3 presents descriptive statistics for the MI-RSWB-S in the total sample ($N = 1{,}011$). FO, HI, HT, and SM were negatively skewed, representing a predominance of high scores, while GR, CO, RSWB-O, and RSWB-R displayed positive skewness, reflecting negative response patterns. Platykurtic distribution was observed for all measures but HI. Normal distribution was assessed by inspecting the absolute values of skewness and kurtosis. All values fell within the acceptable range of ±2 (George and Mallery, 2016).

TABLE 2 | Six-component solution for the Swedish version of the MI-RSWB (MI-RSWB-S).

Principal component	Component loadings		Eigenvalue	% of Variance	Cumulative %
	Range	*M*			
GR	0.80–0.91	0.86	8.91	18.56	18.56
HI	0.44–0.79	0.64	4.45	9.27	27.83
FO	0.47–0.80	0.65	4.31	8.98	36.81
HT	0.18–0.72	0.52	2.78	5.78	42.59
SM	0.22–0.67	0.40	2.72	5.66	48.25
CO	0.03–0.64	0.40	2.35	4.89	53.14

Principal component analysis with VARIMAX rotation with Kaiser normalisation (n = 300). GR, General Religiosity; HI, Hope Immanent; FO, Forgiveness; HT, Hope Transcendent; SM, Experiences of Sense and Meaning; and CO, Connectedness.

Gender and Age Effects

Furthermore, independent t tests were conducted to assess differences between women ($n = 747$) and men ($n = 252$). Women demonstrated significantly higher scores ($M = 35.06$; $SD = 6.53$) than men ($M = 33.54$; $SD = 6.96$) on HI [$t(997) = 3.14$; $p < 0.01$], representing an effect size of $d = 0.22$. Moreover, women scored lower ($M = 33.65$; $SD = 7.57$) than men ($M = 35.44$; $SD = 7.38$) on HT [$t(997) = -3.26$; $p < 0.01$], with an effect size of $d = 0.24$. No other gender effects were observed.

In addition, Pearson's correlations were calculated to examine age effects, revealing positive associations between age and CO ($r = 0.15$; $p < 0.01$), SM ($r = 0.13$; $p < 0.01$), and RSWB-O ($r = 0.08$; $p < 0.05$).

Intercorrelations and Internal Consistencies

As given in **Table 4**, the MI-RSWB-S demonstrated acceptable to excellent internal consistency for all measures except SM, with Cronbach's α coefficients ranging from 0.67 to 0.97. RSWB-O was significantly positively correlated with RSWB-R and all six MI-RSWB-S dimensions, displaying the strongest associations with RSWB-R, GR, and CO. RSWB-R was positively related to all subscales but HT, with the strongest link found for GR. Positive relationships were established between all subscales, with the exception of HT, which was only linked to FO. The strongest subscale intercorrelation was observed between GR and CO, followed by SM and CO ($p < 0.01$).

MI-RSWB-S in Relation to Personality, Sense of Coherence, and Religiosity

Internal consistencies of the TIPI subscales, SOC-13, and CRS-5 as well as associations between these validation instruments and the MI-RSWB-S are given in **Table 5**.

The TIPI subscales Extraversion ($\alpha = 0.75$) and Neuroticism ($\alpha = 0.70$) displayed acceptable levels of internal consistency, whereas Conscientiousness ($\alpha = 0.57$) showed poor internal consistency. The lowest Cronbach's α values were found for Agreeableness ($\alpha = 0.24$) and Openness to Experience ($\alpha = 0.40$).

Extraversion was positively correlated with RSWB-O ($p < 0.01$), RSWB-R ($p < 0.05$), and all MI-RSWB-S dimensions ($p < 0.01$) but GR, to which it was unrelated. Neuroticism was negatively linked to GR ($p < 0.05$), FO, HI, HT, SM, and RSWB-O ($p < 0.01$) and unassociated with CO and RSWB-R, whereas Openness to Experience and Agreeableness were positively related to all MI-RSWB-S measures ($p < 0.01$; $p < 0.05$ for Openness to Experience and HI). Conscientiousness exhibited positive correlations with GR ($p < 0.05$), FO, HI, SM, and RSWB-O ($p < 0.01$). No significant associations were found between Conscientiousness and the other MI-RSWB-S scores. SOC-13 demonstrated good internal consistency, with a Cronbach's α coefficient of 0.83. SOC was positively correlated with FO, HI, HT, SM, and RSWB-O ($p < 0.01$) and unrelated to GR, CO, and RSWB-R. Excellent internal consistency was obtained for CRS-5 ($\alpha = 0.92$). Positive correlations were observed between CRS and all MI-RSWB-S measures ($p < 0.01$) except HT.

DISCUSSION

The main objective of the present work was to provide a validated Swedish version of the MI-RSWB-S. Furthermore, it was intended to investigate how the MI-RSWB-S relates to the Big Five personality traits, SOC, and CRS. Using data from 1,011 Swedish students, a psychometric evaluation of the translated instrument was undertaken. Thereby, we observed acceptable to excellent internal consistency for most of the subscales, with the highest Cronbach's α coefficient found for GR, which mirrors the results of previous research (e.g., Unterrainer et al., 2010, 2012). While the postulated six-component solution of the MI-RSWB received considerable empirical support based on PCA results, CFA of the original factor structure demonstrated poor model fit. On the basis of these findings, together with those of earlier reports and theoretical considerations regarding the conceptualisation of RSWB, a new model was specified. CFA of the suggested structure yielded excellent model fit indices, thereby confirming its construct validity.

In light of these results, we propose a revision of the MI-RSWB structure. Instead of summarising all MI-RSWB dimensions into a total score and using this as an estimate of RSWB (RSWB-O), RSWB may now be obtained by computing the subscales GR and CO (RSWB-R). Subsequently, GR can be calculated for the assessment of RWB, while CO may be used as a measure of Spiritual Well-Being (SWB). Nevertheless, HI, SM, FO, and HT can be analysed independently to gain insight into these specific facets of well-being.

Descriptive analysis of the collected data revealed notable differences in response patterns on the MI-RSWB-S dimensions with a predominance of low GR and CO scores and a preponderance of high values on the other subscales. These findings are markedly different from those obtained among students from other countries (e.g., Unterrainer et al., 2010, 2012; Malinovic et al., 2016), which have demonstrated higher levels of homogeneity within the subscales as well as higher GR and CO mean scores. However, at least, the predominance of low GR values in the current sample is coherent with the notion of Sweden as a relatively secular country (Esmer and Pettersson, 2007) with a small proportion of highly religious

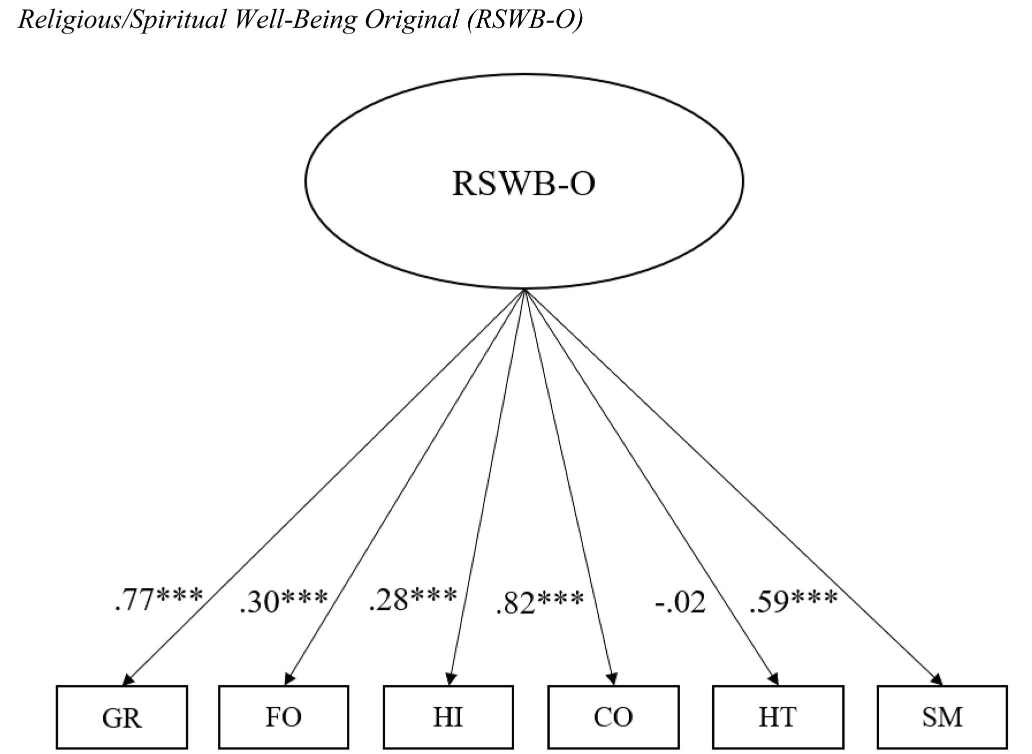

FIGURE 1 | Religious/Spiritual Well-Being (RSWB) Original (RSWB-O). Confirmatory factor analysis ($n = 711$). GR, General Religiosity; FO, Forgiveness; HI, Hope Immanent; CO, Connectedness; HT, Hope Transcendent; SM, Experiences of Sense and Meaning; and RSWB-O, Religious/Spiritual Well-Being Original. $^{***}p < 0.001$.

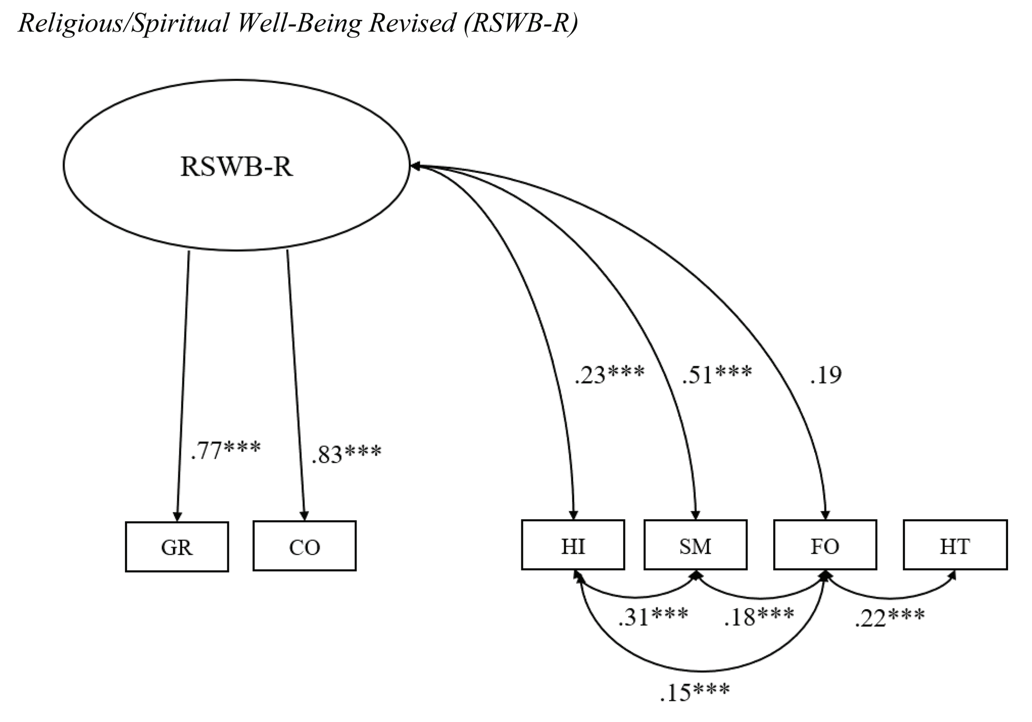

FIGURE 2 | Religious/Spiritual Well-Being Revised (RSWB-R). Confirmatory factor analysis ($n = 711$). GR, General Religiosity; CO, Connectedness; HI, Hope Immanent; SM, Experiences of Sense and Meaning; FO, Forgiveness; HT, Hope Transcendent; and RSWB-R, Religious/Spiritual Well-Being Revised. $^{***}p < 0.001$.

TABLE 3 | Descriptive statistics of the Swedish version of the MI-RSWB (MI-RSWB-S).

	Female (n = 747)		Male (n = 252)		Total (N = 1,011)		min	max	$z_{skewness}$	$z_{kurtosis}$
	M	*SD*	*M*	*SD*	*M*	*SD*				
GR	21.20	14.45	20.09	13.45	20.96	14.19	8	48	9.16	−7.27
FO	34.58	8.48	34.14	8.80	34.50	8.53	10	48	−6.06	−3.49
HI	35.06	6.53	33.54	6.96	34.62	6.66	9	48	−8.19	3.39
CO	22.44	8.96	21.85	9.00	22.36	8.99	8	48	7.79	−2.33
HT	33.65	7.57	35.44	7.38	34.15	7.55	10	48	−5.04	−1.83
SM	36.07	6.19	36.13	6.40	36.08	6.25	15	48	−4.08	−1.77
RSWB-O	182.99	32.66	181.19	32.31	182.67	32.50	98	269	4.77	−3.79
RSWB-R	43.64	21.29	41.94	20.93	43.32	21.21	16	93	7.31	−6.79

N = 1,011. min, minimum score; max, maximum score; GR, General Religiosity; FO, Forgiveness; HI, Hope Immanent; CO, Connectedness; HT, Hope Transcendent; SM, Experiences of Sense and Meaning; RSWB-O, Religious/Spiritual Well-Being Original; and RSWB-R, Religious/Spiritual Well-Being Revised.

TABLE 4 | Intercorrelations and internal consistencies of the MI-RSWB-S.

Dimension	α	GR	FO	HI	CO	HT	SM	RSWB-O	RSWB-R
GR	0.97	-	0.36**	0.20**	0.66**	−0.00	0.41**	0.83**	0.95**
FO	0.85		-	0.13**	0.17**	0.25**	0.18**	0.59**	0.32**
HI	0.81			-	0.21**	0.05	0.33**	0.46**	0.23**
CO	0.81				-	−0.06	0.51**	0.74**	0.86**
HT	0.77					-	−0.04	0.28**	−0.03
SM	0.67						-	0.62**	0.49**
RSWB-O	0.90							-	0.87**
RSWB-R	0.94								-

*N = 1,011. GR, General Religiosity; FO, Forgiveness; HI, Hope Immanent; CO, Connectedness; HT, Hope Transcendent; SM, Experiences of Sense and Meaning; RSWB-O, Religious/Spiritual Well-Being Original; and RSWB-R, Religious/Spiritual Well-Being Revised. **p < 0.01.*

TABLE 5 | MI-RSWB-S in relation to personality traits, Sense of Coherence, and religiosity.

	α	GR	FO	HI	CO	HT	SM	RSWB-O	RSWB-R
Extraversion	0.75	0.04	0.09**	0.24**	0.11**	0.09**	0.18**	0.18**	0.08*
Neuroticism	0.70	−0.07*	−0.17**	−0.30**	−0.03	−0.22**	−0.11**	−0.22**	−0.06
Openness	0.40	0.11**	0.14**	0.08*	0.24**	0.11**	0.24**	0.24**	0.18**
Conscientiousn.	0.57	0.08*	0.13**	0.33**	0.02	0.06	0.10**	0.18**	0.06
Agreeableness	0.24	0.13**	0.29**	0.21**	0.10**	0.09**	0.19**	0.26**	0.13**
SOC	0.83	0.06	0.31**	0.48**	0.00	0.30**	0.14**	0.30**	0.04
CRS	0.92[a]	0.94**[b]	0.36**[b]	0.17**[b]	0.64**[b]	−0.01[b]	0.39**[b]	0.79**[b]	0.90**[b]

N = 1,011. GR, General Religiosity; FO, Forgiveness; HI, Hope Immanent; CO, Connectedness; HT, Hope Transcendent; SM, Experiences of Sense and Meaning; RSWB-O, Religious/Spiritual Well-Being Original; RSWB-R, Religious/Spiritual Well-Being Revised; Neuroticism, reversed Emotional Stability; Openness, Openness to Experience; Conscientiousn, Conscientiousness; SOC, Sense of Coherence; and CRS, Centrality of Religiosity Scale.
[a]*n = 948.*
[b]*n = 1004.*
**p < 0.05; **p < 0.01.*

people (Pew Research Center, 2018a). Furthermore, our findings underline the importance of multidimensional assessment of R/S by suggesting that low levels of GR and CO (and consequently RSWB-R) do not rule out the possibility of extending forgiveness and experiencing hope and sense and meaning. This raises the question of what motivational factors other than those of explicitly R/S kind may be responsible for this observation. For example, a person practicing the virtue of forgiveness may not attribute this behaviour to religious beliefs. Although significant, a meta-analysis (Fehr et al., 2010) investigating the correlates of forgiveness reported a relatively weak positive link between forgiveness and religiosity, especially when compared to dispositional and situational factors such as state empathy and apology. Moreover, an atheist may score high on HT as a result of accepting the mortal nature of human existence, rather than as an expression of confidence in life after death. Similarly, the optimistic expectations and sense of certainty about the future captured by the dimension HI may not stem from religious convictions or spiritual experiences (see e.g., Benzein et al., 2000). In fact, researchers have identified a number of predictors of hope, including life satisfaction, optimism, self-esteem, and social support (Yarcheski and Mahon, 2016).

In line with past research (e.g., Malinovic et al., 2016; Hiebler-Ragger et al., 2018), Extraversion was positively related to HI, SM, and RSWB-O, with the highest correlation found for HI. Contrary to most other observations (e.g., Berger et al., 2016), a significant positive link was identified between Extraversion and HT. Largely consistent with previous reports (e.g., Agarkov et al., 2018; Hiebler-Ragger et al., 2018), HI, HT, and RSWB-O were identified as the strongest negative correlates of Neuroticism. Also in accord with these studies as well as with those investigating the relationship between Neuroticism and spirituality (see Saroglou, 2010), no significant association was found between this personality trait and CO. However, Neuroticism was significantly negatively related to GR and SM, which has not been noted in other studies (e.g., Unterrainer et al., 2012). Openness to Experience was positively linked to all MI-RSWB-S measures. These findings are somewhat different from those of preceding studies, which have generated mixed results. Substantially in agreement with earlier observations (e.g., Malinovic et al., 2016), Conscientiousness exhibited significant positive correlations with GR, FO, HI, SM, and RSWB-O and was unrelated to CO and HT. The strongest link was found between Conscientiousness and HI, partly consistent with previous results (e.g., Stefa-Missagli et al., 2014). In accordance with the findings of Unterrainer et al. (2012), Agreeableness was positively associated with all MI-RSWB-S scores, with the highest correlations observed between Agreeableness and FO (consistent with the findings of Fehr et al., 2010), RSWB-O, and HI. Whereas RSWB-O was significantly related to all of the Big Five traits (thereby exhibiting positive correlations to all measures but Neuroticism), RSWB-R was only (positively) associated with Extraversion, Openness to Experience, and Agreeableness. Moreover, these correlations were weaker than those found for RSWB-O. However, it is important to note that only Extraversion and Neuroticism demonstrated acceptable levels of internal consistency. Consequently, only the reported links between these personality traits and the MI-RSWB-S measures can be interpreted with a relatively high degree of certainty. Conscientiousness, on the other hand, showed poor internal consistency, albeit slightly higher than that reported in Gosling et al. (2003). The Cronbach's α values obtained for Openness to Experience and Agreeableness were lower than those previously observed and indicated unacceptable levels of internal consistency. Nevertheless, the correlations identified between the MI-RSWB-S scores and Conscientiousness, Openness to Experience, and Agreeableness were somewhat similar to those found in other studies.

Sense of Coherence was positively correlated with all MI-RSWB-S measures but GR, CO, and RSWB-R. These results indicate that FO, HI, CO, HT, and SM may be more connected to SOC than R/S in a narrower sense, thus drawing attention to specific aspects of R/S rather than to its conceptual core. The highest correlation was observed between SOC and HI. These findings are substantially in line with those of preceding studies (e.g., Unterrainer et al., 2010; Berger et al., 2016), which have reported positive links between SOC and RSWB-O and all its facets except CO.

Centrality of Religiosity Scale was positively related to all MI-RSWB-S subscales but HT, to which it was unrelated. Mirroring the results reported by Berger et al. (2016), CRS was most strongly associated with GR, followed by CO, SM, FO, and HI. This supports the notion that GR represents a more general measure of religiosity. Furthermore, CRS was more strongly related to RSWB-R than to RSWB-O.

Moreover, small gender effects were identified for immanent and transcendent hope, with women scoring higher than men on HI and lower on HT, as observed by Unterrainer and Fink (2013). Unlike previous studies, which have reported higher levels of RSWB-O, FO (e.g., Unterrainer and Fink, 2013; Stefa-Missagli et al., 2014), GR (e.g., Berger et al., 2016), CO, and SM (e.g., Unterrainer and Fink, 2013) in women than in men (see also Pew Research Center, 2016); no further gender differences were found. In addition, weak correlations were detected between age and CO, SM, and RSWB-O.

LIMITATIONS AND FUTURE PERSPECTIVES

Despite the strengths of the present work (e.g., large sample size, multidimensional assessment of R/S, and exhaustive statistical analysis), several limitations warrant mention. First, our sample comprises Swedish students with a high proportion of female respondents and a significant number of participants under the age of 30. Given these circumstances, this sample cannot be regarded as representative for the Swedish population. Second, the survey was distributed on Facebook and Instagram, consequently excluding students who are not active on these platforms. It should also be noted that this research was conducted during the Covid-19 pandemic, potentially further limiting the generalisability of the findings. Third, in view of the cross-sectional nature of the study, no causal inferences can be drawn. Last, although some of the TIPI subscales demonstrated remarkably low levels of internal consistency, no alternative estimates of reliability (e.g., test-retest reliability correlations) were provided, as suggested by Gosling et al. (2003). To circumvent some of these limitations, future research might use longer personality measures and focus on more representative samples and longitudinal analyses. Studies may also be conducted in clinical populations. In addition to this, qualitative research is encouraged to explore the mechanisms responsible for the observed discrepancies between the subscale scores of the MI-RSWB-S. Considering the positive links between SOC and HI, CO, HT, and SM, future studies may also delve into the reasons and implications of these observations. Moreover, further research might consider examining how the MI-RSWB-S relates to other measures associated with mental health, such as for instance the components of the PERMA model of well-being (Seligman, 2011).

In conclusion, the Swedish adoption of the MI-RSWB demonstrated psychometric properties equivalent to those of the original Austrian–German version. However, CFA favoured a two-factor model over the original six-dimensional structure, resulting in a revision of the inventory. This revised version of the MI-RSWB-S can be regarded as a valid and reliable instrument for assessing RSWB and consequently a valuable contribution to the field of psychology of religion, which may be used to further investigate the relationship between R/S and mental health.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Vienna. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MW and H-FU conceptualised the study. MW and JF acquired the data and performed all statistical analyses. All authors critically discussed the results. The manuscript was drafted by MW and revised by H-FU, JF, and NP-D. GS proof read the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.783761/full#supplementary-material

REFERENCES

Agarkov, V. A., Alexandrov, Y. I., Bronfman, S. A., Chernenko, A. M., Kapfhammer, H.-P., and Unterrainer, H.-F. (2018). A Russian adaptation of the multidimensional inventory for religious/spiritual well-being: psychometric properties for Young adults and associations with personality and psychiatric symptoms. *Arch. Psychol. Relig.* 40, 104–115. doi: 10.1163/15736121-12341347

Antonovsky, A. (1987). *Unraveling the Mystery of Health: How People Manage Stress and Stay Well.* San Francisco, CA: Jossey-Bass.

Antonovsky, A. (1991a). *Hälsans mysterium*. Stockholm: Natur och kultur.

Antonovsky, A. (1991b). "The structural sources of salutogenic strengths," in *Personality and Stress: Individual Differences in the Stress Process*. eds. C. L. Cooper and R. Payne (Chichester: John Wiley & Sons), 67–104.

Benzein, E. G., Saveman, B.-I., and Norberg, A. (2000). The meaning of Hope in healthy, nonreligious swedes. *West. J. Nurs. Res.* 22, 303–319. doi: 10.1177/01939450022044430

Berger, D., Fink, A., Perez Gomez, M. M., Lewis, A., and Unterrainer, H.-F. (2016). The validation of a Spanish version of the multidimensional inventory of religious/spiritual well-being in Mexican college students. *Span. J. Psychol.* 19:E3. doi: 10.1017/sjp.2016.9

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., and Young, S. L. (2018). Best practices for developing and validating scales for health, social, and Behavioral research: A primer. *Front. Public Health* 6:149. doi: 10.3389/fpubh.2018.00149

Costa, P. T., and McCrae, R. R. (1992a). Normal personality assessment in clinical practice: The NEO personality inventory. *Psychol. Assess.* 4, 5–13. doi: 10.1037/1040-3590.4.1.5

Costa, P. T., and McCrae, R. R. (1992b). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional Manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Dadfar, M., Lester, D., Turan, Y., Beshai, J. A., and Unterrainer, H.-F. (2019). Validation of the multidimensional inventory for religious spiritual well-being with Iranian samples. *Ment. Health Relig. Cult.* 22, 591–601. doi: 10.1080/13674676.2019.1628194

Ellison, C. W. (1983). Spiritual well-being: conceptualization and measurement. *J. Psychol. Theol.* 11, 330–338. doi: 10.1177/009164718301100406

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science* 196, 129–136. doi: 10.1126/science.847460

Eriksson, M., and Lindström, B. (2005). Validity of Antonovsky's sense of coherence scale: A systematic review. *J. Epidemiol. Community Health* 59, 460–466. doi: 10.1136/jech.2003.018085

Esmer, Y., and Pettersson, T. (eds.) (2007). *Measuring and Mapping Cultures: 25 Years of Comparative Value Surveys*. Leiden: Brill.

Fehr, R., Gelfand, M. J., and Nag, M. (2010). The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychol. Bull.* 136, 894–914. doi: 10.1037/a0019993

George, D., and Mallery, P. (2016). *IBM SPSS Statistics 23 Step by Step: A Simple Guide and Reference*, 14th Edn. New York, NY: Routledge.

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychol. Assess.* 4, 26–42. doi: 10.1037/1040-3590.4.1.26

Gosling, S. D., Rentfrow, P. J., and Swann, W. B. (2003). A very brief measure of the big-five personality domains. *J. Res. Pers.* 37, 504–528. doi: 10.1016/S0092-6566(03)00046-1

Hiebler-Ragger, M., Fuchshuber, J., Dröscher, H., Vajda, C., Fink, A., and Unterrainer, H. -F. (2018). Personality influences the relationship between primary emotions and religious/spiritual well-being. *Front. Psychol.* 9:370. doi: 10.3389/fpsyg.2018.00370

Hodapp, B., and Zwingmann, C. (2019). Religiosity/spirituality and mental health: A meta-analysis of studies from the German-speaking area. *J. Relig. Health.* 58, 1970–1998. doi: 10.1007/s10943-019-00781-2

Huber, S., and Huber, O. W. (2012). The centrality of religiosity scale (CRS). *Religion* 3, 710–724. doi: 10.3390/rel3030710

John, O. P., and Srivastava, S. (1999). "The big five trait taxonomy: history, measurement, and theoretical perspectives," in *Handbook of Personality: Theory and Research*. eds. L. A. Pervin and O. P. John (New York, NY: Guilford Press), 102–138.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika* 39, 31–36. doi: 10.1007/BF02291575

Kaiser, H. F., and Rice, J. (1974). Little jiffy, mark IV. *Educ. Psychol. Meas.* 34, 111–117. doi: 10.1177/001316447403400115

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling*, 4th Edn. New York, NY: Guilford Press.

Koenig, H. G. (2008). Concerns about measuring "spirituality" in research. *J. Nerv. Ment. Dis.* 196, 349–355. doi: 10.1097/NMD.0b013e31816ff796

Koenig, H. G., King, D. E., and Carson, V. B. (2012). *Handbook of Religion and Health*, 2nd Edn. New York, NY: Oxford University Press.

Kyrkan, S. (2021). Svenska kyrkan i siffror. Available at: https://www.svenskakyrkan.se/statistik (Accessed September 5, 2021).

Larsson, G., and Kallenberg, K. (1999). Dimensional analysis of sense of coherence using structural equation modelling. *Eur. J. Personal.* 13, 51–61. doi: 10.1002/(SICI)1099-0984(199901/02)13:1<51::AID-PER321>3.0.CO;2-P

Lundell, E. (2014). Ten-Item Personality Inventory-(TIPI) – Swedish translation. Available at: http://gosling.psy.utexas.edu/wp-content/uploads/2014/09/TIPISwedishtranslation.doc (Accessed September 5, 2021).

Malinovic, A., Fink, A., Lewis, A. J., and Unterrainer, H. -F. (2016). Dimensions of religious/spiritual well-being in relation to personality and stress coping:

initial results from bosnian young adults. *J. Spiritual. Ment. Health* 18, 43–54. doi: 10.1080/19349637.2015.1059301

Mittelmark, M. B., Sagy, S., Eriksson, M., Bauer, G. F., Pelikan, J. M., Lindström, B., et al. (eds.) (2017). *The Handbook of Salutogenesis*. Cham: Springer.

Moberg, D. O. (1971). *Spiritual Well-Being: Background and Issues. White House Conference on Aging*. Washington, DC: U.S. Government Printing Office.

Pew Research Center. (2016). The Gender Gap in Religion Around the World. Available at: https://www.pewforum.org/wp-content/uploads/sites/7/2016/03/Religion-and-Gender-Full-Report.pdf (Accessed September 5, 2021).

Pew Research Center. (2018a). Eastern and Western Europeans Differ on Importance of Religion, Views of Minorities, and Key Social Issues. Available at: https://www.pewforum.org/wp-content/uploads/sites/7/2018/10/Eastern-Western-Europe-FOR-WEB.pdf (Accessed September 5, 2021).

Pew Research Center. (2018b). The Age Gap in Religion Around the World. Available at: https://www.pewforum.org/wp-content/uploads/sites/7/2018/06/ReligiousCommitment-FULL-WEB.pdf (Accessed September 5, 2021).

Räty, L. K. A., Larsson, B. M. W., and Söderfeldt, B. A. (2003). Health-related quality of life in youth: A comparison between adolescents and young adults with uncomplicated epilepsy and healthy controls. *J. Adolesc. Health* 33, 252–258. doi: 10.1016/S1054-139X(03)00101-0

Saroglou, V. (2010). Religiousness as a cultural adaptation of basic traits: a five-factor model perspective. *Personal. Soc. Psychol. Rev.* 14, 108–125. doi: 10.1177/1088868309352322

Seligman, M. E. P. (2011). *Flourish: A Visionary New Understanding of Happiness and Well-Being*. New York, NY: Free Press.

Sjöborg, A. (2014). "CRS-5 SWE," in *Secular and Sacred?: The Scandinavian Case of Religion in Human Rights, Law and Public Space*. (eds.) Breemer R. van den, J. Casanova and T. Wyller (Göttingen: Vandenhoeck & Ruprecht), 236-260.

Stefa-Missagli, S., Huber, H. P., Fink, A., Sarlo, M., and Unterrainer, H. -F. (2014). Dimensions of religious/spiritual well-being, personality, and mental health: initial results from italian college students. *Arch. Psychol. Relig.* 36, 368–385. doi: 10.1163/15736121-12341290

Thurfjell, D. (2019). *Det gudlösa folket: De postkristna svenskarna och religionen*. Stockholm: Nordstedts.

Unterrainer, H. -F. (2021). The multidimensional measurement of religious/spiritual well-being: recent developments in scale validation and clinical applications. submitted.

Unterrainer, H. -F., and Fink, A. (2013). Das Multidimensionale Inventar zum religiös-spirituellen Befinden (MI-RSB): Normwerte für die österreichische Allgemeinbevölkerung. *Diagnostica* 59, 33–44. doi: 10.1026/0012-1924/a000077

Unterrainer, H. -F., Huber, H. -P., Ladenhauf, K. H., Wallner-Liebmann, S. J., and Liebmann, P. M. (2010). MI-RSB 48: Die Entwicklung eines multidimensionalen Inventars zum religiös-spirituellen Befinden. *Diagnostica* 56, 82–93. doi: 10.1026/0012-1924/a000001

Unterrainer, H. -F., Ladenhauf, K. H., Wallner-Liebmann, S. J., and Fink, A. (2011). Different types of religious/spiritual well-being in relation to personality and subjective well-being. *Int. J. Psychol. Relig.* 21, 115–126. doi: 10.1080/10508619.2011.557003

Unterrainer, H. -F., Nelson, O., Collicutt, J., and Fink, A. (2012). The English version of the multidimensional inventory for religious/spiritual well-being (MI-RSWB-E): first results from british college students. *Religion* 3, 588–599. doi: 10.3390/rel3030588

Yarcheski, A., and Mahon, N. E. (2016). Meta-analyses of predictors of hope in adolescents. *West. J. Nurs. Res.* 38, 345–368. doi: 10.1177/0193945914559545

Zinnbauer, B. J., Pargament, K. I., Cole, B., Rye, M. S., Butter, E. M., Belavich, T. G., et al. (1997). Religion and spirituality: Unfuzzying the fuzzy. *J. Sci. Study Relig.* 36, 549–564. doi: 10.2307/1387689

Zuckerman, P. (2020). *Society without God. What the Least Religious Nations Can Tell us about Contentment*, *2nd edn*. New York, NY: New York University Press.

Conflict of Interest: The reviewer HB-C declared a shared affiliation, with one of the authors H-FU to the handling editor at the time of the review.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 November 2021
doi: 10.3389/fpsyg.2021.736514

Factor Structure and Psychometric Properties of the Family Communication Scale in the Chinese Population

Ningyuan Guo[1], *Henry C. Y. Ho*[2], *Man Ping Wang*[1]*, *Agnes Y. Lai*[1], *Tzu Tsun Luk*[1], *Kasisomayajula Viswanath*[3,4], *Sophia S. Chan*[1] *and Tai Hing Lam*[5]

[1] School of Nursing, University of Hong Kong, Hong Kong, Hong Kong SAR, China, [2] Department of Psychology and Centre for Psychosocial Health, The Education University of Hong Kong, Hong Kong, Hong Kong SAR, China, [3] Department of Social and Behavioral Sciences, Harvard T. H. Chan School of Public Health, Boston, MA, United States, [4] Center for Community-Based Research, Dana-Farber Cancer Institute, Boston, MA, United States, [5] School of Public Health, University of Hong Kong, Hong Kong, Hong Kong SAR, China

OPEN ACCESS

Edited by:
Antonio De Padua Serafim,
Universidade de São Paulo, Brazil

Reviewed by:
Carmen Moret-Tatay,
Catholic University of Valencia
San Vicente Mártir, Spain
Sai-fu Fung,
City University of Hong Kong,
Hong Kong SAR, China

***Correspondence:**
Man Ping Wang
mpwang@hku.hk

Specialty section:
This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 05 July 2021
Accepted: 13 October 2021
Published: 12 November 2021

Citation:
Guo N, Ho HCY, Wang MP,
Lai AY, Luk TT, Viswanath K, Chan SS
and Lam TH (2021) Factor Structure
and Psychometric Properties of the
Family Communication Scale
in the Chinese Population.
Front. Psychol. 12:736514.
doi: 10.3389/fpsyg.2021.736514

Purpose: To evaluate the factor structure and psychometric properties of the 10-item Family Communication Scale (FCS) in the Chinese population.

Methods: Study 1 was a population-based survey [N = 687, 61.1% female; mean age (SD) 56.6 (19.1)]. Study 2 was a community-based intervention (N = 1983, 76.7% female; 57.8% aged 20–59 years). We conducted exploratory factor analysis (EFA) in Study 1 and replicated the model by confirmatory factor analysis (CFA) in Study 2. Psychometric properties were evaluated, including internal consistency, test–retest reliability, convergent and discriminant validity, and known-group validity. We identified how the FCS scores differed by sociodemographic characteristics and communication methods including face to face and Information and Communication Technologies (ICTs) in Study 1.

Results: The EFA and CFA supported a one-factor structure. The Chinese FCS showed a good internal consistency (Cronbach's alpha = 0.91; McDonald's Omega = 0.91) and was stable over 1-month (intraclass correlation coefficient = 0.69, $P < 0.001$). Convergent validity was supported by positive correlations of FCS with the Subjective Happiness Scale, Family Adaption, Partnership, Growth, Affection, Resolve (APGAR) Scale, family health, harmony, and happiness, and perceived family communication sufficiency and quality (All $P < 0.001$). Discriminant validity was supported by the stronger correlation of FCS with Short Form-12 Health Survey Version 2 Mental Component than that with Physical Component ($P < 0.001$). Higher household income, frequent face-to-face communication, and frequent use of phone calls, instant messaging, and social networking sites were associated with higher FCS scores.

Conclusion: The one-factor structure of the Chinese FCS can be a reliable and valid measurement of positive family communication, in the context of ICT integration into family communication.

Clinical Trial Registration: [www.ClinicalTrials.gov], identifier [NCT02563613].

Keywords: family communication scale, positive family communication, communication method, information and communication technologies, validation, Chinese

INTRODUCTION

Family communication is the act of sharing ideas, participating in decision making, and expressing feelings among members as a family unit (Olson, 2000). Less family communication or more family conflicts were associated with higher risks of behavioral problems such as substance use disorders and gaming disorders in young people (Challier et al., 2000; Schneider et al., 2017). In contrast, positive family communication, including aspects of listening, speaking, self-disclosure, clarity, continuity tracking, and respect may improve physical and mental health (Olson, 2000), through social support and adaptive coping strategies with stressors (Schrodt et al., 2008). Positive family communication also facilitates a balanced level of family flexibility to change family rules and cohesion of emotional bonding (Olson, 2000). These benefits on families were shown in our previous interventions indicating improved family well-being through enhancing family communication (Ho et al., 2017, 2018; Shen et al., 2017a).

The Family Communication Scale (FCS) is a widely used measurement of the satisfaction toward the aspects of positive communication among family members (Olson and Barnes, 2004), which was adapted from the 20-item Patient-Adolescent Communication Scale (PAC) measuring communication in families with adolescents (Barnes and Olson, 1985). Compared with PAC, the shorter 10-item FCS has a lighter operation burden on respondents and can be used in broader family forms and families at various life cycle stages (Olson and Barnes, 2004). The FCS has been widely used globally with consistent satisfactory reliability and validity, but only a few have reported the factor structure (Olson and Barnes, 2004; Baiocco et al., 2013; Koutra et al., 2013; Gomes et al., 2017; Martínez-Pampliega et al., 2017). A validation study in Turkey showed a one-factor structure that discarded items on self-disclosure and affective communication (Türkdoğan et al., 2018), which was controversial with the original scale having all 10 items in one factor (Olson and Barnes, 2004). Such a variable FCS scoring structure can be explained by cultural differences in family communication patterns across different populations. Unlike an expression of self-emphasized in the West, implicit communication and listening-centeredness are often used in the Asian collectivist culture (Bond, 2010). Apart from cultural differences, our previous qualitative studies in Chinese showed that family communication could be affected by interaction time, income, and psychosocial capitals (Chan et al., 2011; Lam et al., 2012).

The evolving Information and Communication Technologies [ICTs; e.g., mobile phone, instant messaging (IM), social networking sites (SNS)] have transformed communication patterns (Carvalho et al., 2015). Family communication can be conducted in real-time and/or asynchronously using ICTs, which may overcome time and distance barriers. ICTs have enabled transnational family communication for low-income immigrant families to maintain virtual intimacy, emotional support, and transnational caregiving in a qualitative interview (Gonzalez and Katz, 2016). Both factual and emotional information can be exchanged among family members through multimedia on ICTs such as texts, pictures, audio clips, and videos (Carvalho et al., 2015). Higher levels of family well-being have been observed in people who frequently used ICTs for family communication such as phone calls and video calls in our previous population-based studies (Wang et al., 2015; Shen et al., 2017b). Instruments such as Mobile Device Proficiency Questionnaire, Computer Proficiency Questionnaire (Moret-Tatay et al., 2019), and ICT competence scale (Aesaert et al., 2014) were developed for measuring ICT use and showed cross-cultural differences. Direct measures of family communication using ICTs are lacking particularly in the Chinese population.

The study aimed to evaluate the Chinese version of FCS in a population-based telephone survey sample and a community-based randomized controlled trial sample of Hong Kong Chinese. The factor structure and psychometric properties of FCS have yet to be examined in the Chinese population, compared with other validated instruments such as Family Adaption, Partnership, Growth, Affection, Resolve [(APGAR) Scale; Chan et al., 1988], the single item of family happiness in Family Well-being Scale (Shen et al., 2019). We also took advantage of the representative survey sample to identify how FCS scores differed by sociodemographic characteristics and family communication methods, including ICTs and face to face.

MATERIALS AND METHODS

Study Design

Study 1: The Hong Kong Family and Health Information Trends Survey

The FHInTS is a periodic territory-wide telephone survey on information use, health communication, and family well-being among Hong Kong residents aged 18 years or above. We have conducted five waves of FHInTS since 2009 and reported details of the study design elsewhere (Wang et al., 2015; Shen et al., 2017b). Study 1 is part of the fifth wave of FHInTS, conducted from February to August 2017.

Study 2: Happy Family Kitchen Movement Project

The HFKM was a community-based intervention program conducted from January 2015 to July 2017 in Hong Kong residents aged 12 years or above to improve family well-being using the positive psychology framework integrated with physical and psychosocial health. Details of the study design and sociodemographic characteristics of the participants were reported elsewhere (Ho et al., 2019). The study was registered with ClinicalTrials.gov (NCT02563613).

Participants

Study 1: The Hong Kong Family and Health Information Trends Survey

We used a two-stage probability-based sampling procedure. First, landline telephone numbers were randomly generated using known prefixes assigned to telecommunication service providers under the Numbering Plan provided by the Government Office of the Communications Authority. Invalid numbers were removed according to the computer and manual dialing records. Telephone numbers of respondents from previous waves were filtered. Second, once a household was successfully reached, an eligible family member with the soonest next birthday was invited to the survey. All telephone interviews were conducted by trained interviewers from the Public Opinion Program at the University of Hong Kong, a reputable local survey agency. Among 5,773 invited respondents, 4,054 were successfully interviewed (response rate = 70.2%). A randomly selected subset of 687 (17.0%) completed the Chinese version of FCS [61.1% female; mean age [standard deviation (SD)] 56.6 (19.1) years; 42.8% had secondary educational attainment]. We evaluated the factor structure by exploratory factor analysis (EFA), internal consistency reliability, and construct validity of the Chinese version of FCS. We also examined associations of sociodemographic characteristics and family communication methods with positive family communication.

Study 2: Happy Family Kitchen Movement Project

A total of 54 social service units and schools collaborated with the research team to design and implement the trial in 1,983 participants (76.7% female; 57.8% aged 20–59 years; 52.3% had secondary educational attainment) from 1,467 families in all 18 districts in Hong Kong. The social service units and schools were randomly allocated as clusters with the participants they recruited into three groups. Positive Physical Activity group (PPA; intervention arm 1) and Positive Healthy Diet group (PHD; intervention arm 2) received a core session of about 2 h, followed by a booster session of about 1 h a month later. The control group (the waitlist control arm) received a tea gathering session at the beginning and a month later. The core session in the PPA group included group activities and homework assignments focusing on positive psychology and physical activity. The core session in the PHD group focused on positive psychology and healthy diet. The booster session in the PPA and PHD groups focused on consolidating knowledge, skills, and experience gained from the core sessions. The tea gathering sessions in the control group included activities unrelated to PPA/PHD. The participants completed assessment questionnaires at four time points: baseline (T1), immediately post-intervention (T2), 1-month (T3) follow-up, and 1-month follow-up (T4). We conducted a confirmatory factor analysis (CFA) to determine the replicability of the EFA results in Study 1. We also evaluated the 1-month test–retest reliability and construct validity of the Chinese version of FCS.

Measurements

The translation process of FCS followed the guidelines provided by the author of the original FCS (Olson and Barnes, 2004). A translation team was created and comprised of professional translators who are bilingual in English and Chinese. The FCS was first translated into traditional Chinese and then back-translated into English until a consensus was achieved. Examples of FCS items are "Family members are satisfied with how they communicate with each other," "Family members are very good listeners," and "Family members express affection to each other" (Olson and Barnes, 2004). Each item scores on a five-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree. A higher total score (range 10–50) indicates a greater level of positive family communication.

We examined the construct validity of FCS using the following measurements. In Study 1 and Study 2, Subjective Happiness Scale (SHS; range 1–7; Cronbach's alpha 0.75 in Study 1 and ranged 0.72–0.75 in Study 2) has four items to measure individual happiness (Lyubomirsky and Lepper, 1999; Nan et al., 2014). Family Well-being Scale has three single items (each range 0–10) on family health, harmony, and happiness that were developed specifically in Chinese culture (Chan et al., 2011; Lam et al., 2012). The single item of family happiness has been validated in Hong Kong (Shen et al., 2019). Different measurements between Study 1 and Study 2 can broaden the scope of examination and reduce questionnaire length and response burden. We used measurements that were only included in Study 1 for additionally examining the convergent validity of FCS: Family APGAR Scale (range 0–10; Cronbach's alpha 0.86) has five items to measure the family functioning (Smilkstein, 1978; Chan et al., 1988). Respondents were asked whether they had sufficient communication with family members on a five-point scale ranging from 1 = very insufficient, 2 = insufficient, 3 = fair, 4 = sufficient, to 5 = very sufficient. Perceived family communication quality was rated on an 11-point scale, where 0 = very poor, 5 = half-half, and 10 = very good. Short Form-12 Health Survey Version 2 (SF-12) only included in Study 2 was used for examining discriminative validity. SF-12 has 12 items to measure physical and mental health-related quality of life (HRQoL; Cronbach's alpha ranged 0.79–0.81 for PCS and 0.75–0.76 for MCS) (Ware et al., 1996; Lam et al., 2005). The raw scores are transformed into the Physical Component Subscale (PCS; range 0–100) and Mental Component Subscale (MCS; range 0–100).

Respondents in Study 1 were asked as to how often they used the following methods to communicate/chat with family members, including face to face, phone calls, IM (e.g., WhatsApp), SNS (e.g., Facebook), video calls (e.g., Skype, FaceTime, WeChat video call), and email. Responses included often, sometimes, seldom, and never. We dichotomized the frequency as never/seldom (reference) vs. sometimes/often due to the skew distribution of the continuous variable.

Sociodemographic characteristics included gender, age, marital status, employment status, educational attainment, and monthly household income.

Statistical Analysis

Study 1: The Hong Kong Family and Health Information Trends Survey

All data were weighted by gender, age, and educational attainment distribution of the Hong Kong general population using the random iterative method (Izrael et al., 2004). Missing data were handled by the available case analyses as there were minimal missing values for all variables (<2.5%). Preliminary analyses were conducted to ensure the appropriateness for an EFA: (1) the normality of FCS score was confirmed with a skewness value of −0.65 ($\leq |2.0|$) and a kurtosis value of 4.19 ($\leq |7.0|$), given the large sample size (Kim, 2013); (2) flooring and ceiling effects were not present, with 0.15 and 4.62% (both ≤ 15%) respondents had the lowest or highest possible FCS score (Terwee et al., 2007); (3) the Kaiser–Meyer–Olkin measure of sampling adequacy was of 0.935 and the Bartlett test of sphericity reached a significant level ($\chi^2 = 3272.453$, df = 45, $P < 0.001$), indicating the strong correlations among FCS items for an EFA.

The EFA extracted factors from the 10 items of FCS using the principal factor method with promax rotation (i.e., an oblique rotation that allows for correlations between factors). The factor structure was determined by multiple approaches: Kaiser's criterion (eigenvalues >1), scree plot, parallel analysis with principal components and 10,000 random datasets (the larger the number, the more accurate the estimate; Dinno, 2009); and the minimum average partial test (Courtney and Gordon, 2013). Parallel analysis and the minimum average partial test have been suggested to be the most accurate of all the approaches (Velicer et al., 2000), and consistent results would increase the confidence of the factor structure of FCS. Factors were retained with the adjusting eigenvalues (accounting for sampling bias) greater than that could be generated from random data in parallel analysis and with the minimum average squared partial correlations. Factor loadings were evaluated by the following criteria: ≥ 0.71 excellent, 0.63–0.70 very good, 0.55–0.62 good, 0.45–0.54 fair, 0.32–0.44 poor, and < 0.32 unacceptable (Comrey and Lee, 2013).

Internal consistency reliability was determined by Cronbach's alpha and McDonald's Omega coefficient, requiring value of ≥ 0.70 acceptable, ≥ 0.80 good (Terwee et al., 2007; Hayes and Coutts, 2020). The convergent validity was determined by the correlations of FCS score with scores of SHS, family health, family harmony, family happiness, family APGAR, and perceived sufficiency, and quality of family communication. Pearson's correlation coefficients (r) were calculated, and values of $|r|$ were evaluated by the following criteria: 0.68–1 strong, 0.36–0.67 moderate, and 0–0.35 weak (Taylor, 1990). Differences in r were assessed using Fisher z-transformation test. Known-group validity was evaluated by comparing the mean FCS scores by sociodemographic characteristics and communication methods using linear regression analyses. Multivariable analyses were used to test whether the differences can still present after mutual adjustments. We hypothesized higher FCS scores observed in people with higher household income based on similar associations reported in our qualitative interviews (Chan et al., 2011; Lam et al., 2012). Frequent face to face and ICTs use for family communication have been associated with improved family well-being, which may increase FCS scores (Wang et al., 2015; Shen et al., 2017b). We accordingly hypothesized higher FCS scores observed in people having frequent family communication through face to face and ICTs. Analyses were conducted on Stata 15.0. A P-value of < 0.05 was considered statistically significant.

Study 2: Happy Family Kitchen Movement Project

The principle of intention-to-treat analysis was adopted. CFA with diagonally weighted least squares estimation for ordinal data was performed to examine the factor structure identified by EFA in Study 1 (Li, 2016). The adequacy of model fit was determined by a combination of the following indices: relative/normed Chi-square statistic (χ^2/df, < 3), goodness-of-fit index (GFI; ≥ 0.95), comparative fit index (CFI; ≥ 0.90), root mean square error of approximation (RMSEA; < 0.06), root mean square residual (RMR; < 0.08), and standardized RMR (SRMR; < 0.08) (Hooper et al., 2008; Kline, 2015). We reported results of Chi-square test for descriptive purpose but not for evaluating the model fit (cutoff for good fit: $P > 0.05$), because the result is always statistically significant in large samples (Hooper et al., 2008).

One-month test–retest reliability was determined by intraclass correlation coefficients in the control group, calculated based on a consistency two-way mixed-effects model, by the following criteria: 0.90–1 excellent, 0.75–0.89 good, 0.50–0.74 moderate, and 0–0.49 poor (Koo and Li, 2016). Convergent validity was determined by correlations of FCS score with SF-12 PCS and MCS, SHS, family health, family harmony, and family happiness. Discriminant validity was determined by differences between the correlation of FCS score with SF-12 PCS and that with MCS. Pearson's correlation coefficients (r) were calculated, and the values of $|r|$ were evaluated by the following criteria: 0.68–1 strong, 0.36–0.67 moderate, and 0–0.35 weak (Taylor, 1990). Partial correlation analysis was used to account for the intervention effect at T3 and T4. Differences in r were assessed using Fisher z-transformation test. Analyses were conducted on SPSS 25.0 except for CFA on LISREL 11. A P-value of < 0.05 was considered statistically significant.

RESULTS

Factor Structure and Psychometric Properties of Chinese Version of Family Communication Scale

In the population-based sample of Study 1, EFA showed that a one-factor structure should be retained: (1) Only the first factor (eigenvalue = 4.97) met the Kaiser's criterion and explained 86.72% of the variance; (2) visual examination of screen plot indicated that the first factor accounted for most of the variance (the dashed line in **Figure 1**); (3) parallel analysis indicated that the adjusted eigenvalue (the solid line in **Figure 1**) of the first factor was greater than that could be obtained in random data (the dotted line in **Figure 1**); (4) the minimum average partial test indicated the minimum average squared partial correlations of the first factor (0.021 in **Table 1**). All 10 items had good-to-excellent factor loadings (range 0.55–0.79) (**Table 2**). The internal consistency was good (Cronbach's alpha = 0.91; McDonald's Omega = 0.91). Removal of any item yielded a Cronbach's alpha ranging 0.89–0.91.

CFA was performed on the one-factor structure in the community-based sample of Study 2. All model fit indices were within the prespecified cut-off values (GFI = 0.998 > 0.95, CFI = 0.994 > 0.90, RMSEA = 0.044 < 0.06, RMR = 0.016 < 0.08, SRMR = 0.028 < 0.08), except for χ^2/df = 4.61 > 3 (**Table 3**). No further modifications were made and the final model is presented in **Figure 2**.

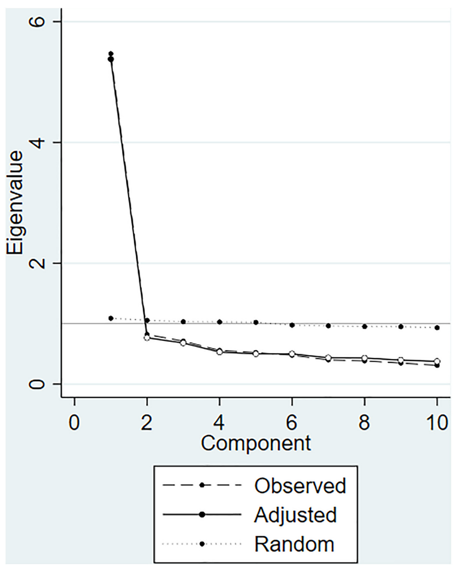

FIGURE 1 | Parallel analysis for determining the number of factors to retain for the exploratory factor analysis in the population-based sample (*N* = 687).

TABLE 1 | The minimum average partial test for determining the number of factors to retain for the exploratory factor analysis in the population-based sample (*N* = 687).

Number of factors	Average squared partial correlations
0	0.25
1	0.021
2	0.039
3	0.057
4	0.093
5	0.12
6	0.19
7	0.29
8	0.46
9	1

The intraclass correlation coefficient for test–retest reliability over 1 month was 0.69 ($P < 0.001$) in the community-based sample of Study 2. The FCS score was positively and moderately correlated with the scores of SHS, family health, family harmony, family happiness, family APGAR, and perceived family communication sufficiency and quality (r range 0.40–0.60; all $P < 0.001$) in Study 1 (**Table 4**). The correlation of the FCS score with perceived family communication quality (r = 0.60) was significantly stronger than that with perceived family communication sufficiency (r = 0.40) ($P < 0.001$). Family communication was also positively and moderately correlated with scores of SHS, family health, family harmony, and family happiness at baseline (T1), 1-month follow-up (T3), and 3-month follow-up (T4) (all $P < 0.001$), regardless of the intervention effects in Study 2. The correlations of the FCS score with SF-12 MCS (r range 0.31–0.34) were significantly stronger than those with PCS (r range 0.13–0.19) at all three time points ($P < 0.001$), regardless of the intervention effects.

Associations of Sociodemographic Characteristics With Positive Family Communication

The mean FCS score (SD) was 37.8 (6.2) in the population-based sample (**Table 4**). Multivariable linear regression analyses showed that housekeepers had higher FCS scores (adjusted β = 3.57, 95% CI 0.23, 6.92), adjusting for other sociodemographic characteristics. Monthly household income was positively associated with FCS scores (P for trend = 0.045) (**Table 5**).

Associations of Communication Method With Positive Family Communication

The most frequent method of family communication was face to face (92.0%), followed by phone calls (66.7%) and IM (59.4%)

TABLE 2 | Descriptive statistics, factor loadings, and internal consistency of the Chinese version of the Family Communication Scale in the population-based sample (*N* = 687).

FCS item[a]	Mean score (SD)[b]	Factor loading[c]	Corrected item-total correlation	Cronbach's alpha if item deleted[d]
1	3.77 (0.89)	0.70	0.67	0.89
2	3.68 (0.96)	0.70	0.66	0.90
3	3.93 (0.85)	0.75	0.72	0.89
4	3.85 (0.90)	0.70	0.66	0.90
5	3.76 (0.92)	0.74	0.70	0.89
6	3.82 (0.85)	0.75	0.72	0.89
7	3.90 (0.79)	0.68	0.64	0.90
8	3.79 (0.84)	0.79	0.75	0.89
9	3.41 (1.03)	0.55	0.53	0.91
10	3.84 (0.86)	0.64	0.60	0.90

[a]*Each item scores from 1 = "strongly disagree" to 5 = "strongly agree."*
[b]*Weighted by age, gender, and educational attainment distribution of the Hong Kong general population.*
[c]*Proportion of total variance = 86.72%.*
[d]*Overall Cronbach's alpha = 0.91; McDonald's Omega = 0.91.*

TABLE 3 | Fit statistics for the one-factor model of the Chinese version of the Family Communication Scale in the community-based sample (N = 1,983).

Model fit indices with cutoff [36,37]	χ^2 (P > 0.05)	Df	χ^2/df	GFI ≥ 0.95	CFI ≥ 0.90	RMSEA < 0.06	RMR < 0.08	SRMR < 0.08
With one factor and 10 test items [a]	161.22 (<0.001)	35	4.61	0.998	0.994	0.044 (0.048, 0.090)	0.016	0.028

χ^2, Chi-square; df, degrees of freedom; CFI, comparative fit index; RMSEA, root mean square standard error of approximation; RMR, root mean square residual; SRMR, standardized root mean square residual.

[a] The one-factor structure was based on the result of exploratory factor analysis.

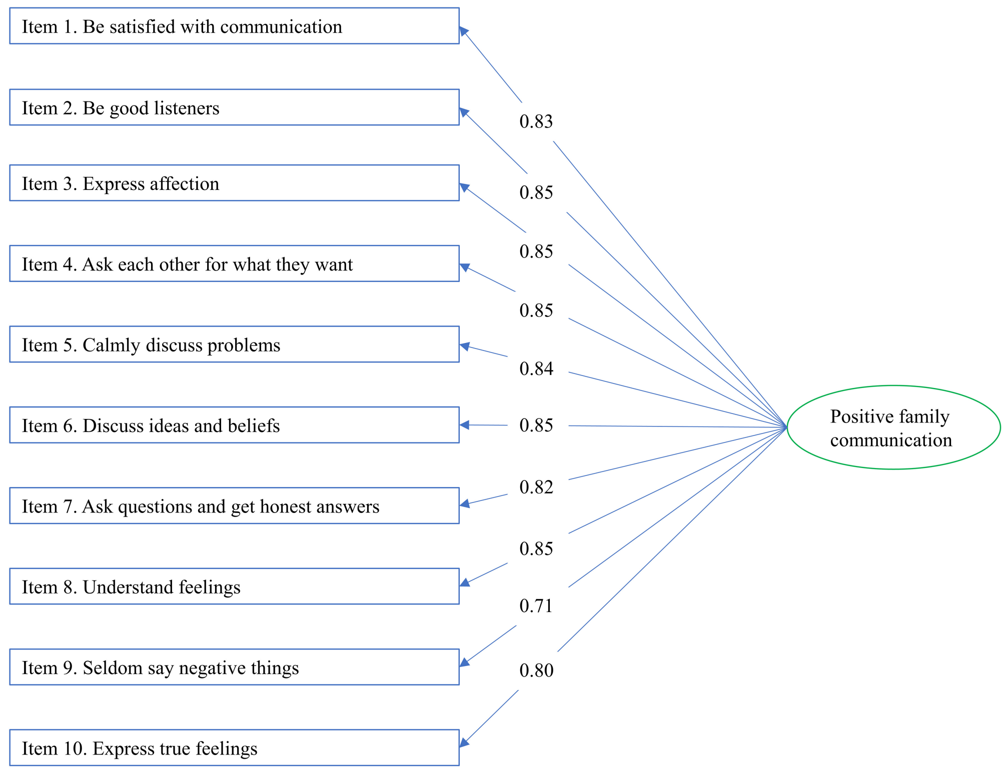

FIGURE 2 | The final one-factor model of the Family Communication Scale indicated by the confirmatory factor analysis with standardized path coefficients in the community-based sample (N = 1983).

in the population-based sample (**Table 6**). Frequent face-to-face communication was strongly associated with higher FCS scores (adjusted β = 6.33, 95% CI 3.71, 8.96). Frequent use of ICTs including phone calls (adjusted β = 2.74, 95% CI 1.50, 3.97), SNS (adjusted β = 2.26, 95% CI 0.55, 3.96), and IM (adjusted β = 1.91, 95% CI 0.53, 3.29) were associated with higher FCS scores. Among the respondents who frequently conducted face-to-face family communication, higher FCS scores were observed for frequent use of phone calls (adjusted β = 1.99, 95% CI 0.82, 3.15) and SNS (adjusted β = 1.84, 95% CI 0.13, 3.57).

DISCUSSION

The EFA showed a one-factor model of the Chinese version of FCS that comprises all 10 items in the population-based sample. The model was replicated by CFA in the community-based sample. Results of relative/normed Chi-square statistic (χ^2/df = 4.61) was higher than the cutoff of 3 (Kline, 2015). However, various cutoffs of χ^2/df ranging 2–5 have been used in the literature, and no consensus was found (Tabachnick et al., 2007). Other model fit indices including GFI, CFI, RMSEA, RMR, and SRMR were within the prespecified cutoff values. Taken together, the one-factor model suggested was acceptable. The one-factor structure was also identified in the original scale (Olson and Barnes, 2004) and the Portuguese validation (Gomes et al., 2017). In contrast, the Spanish validation in the Chilean population showed a two-factor structure, suggesting the independence of emotional/affective dimension of family communication and the other dimension related to more general communication skills, such as problem-solving skills and listening skills (Rivadeneira and López, 2017). The Turkish validation showed a one-factor structure but discarded items on self-disclosure and affective communication because of the low factor loadings and the tendency to be under another dimension (Türkdoğan et al., 2018). This can be a reflection of listening centeredness and implicit communication style valued in the collectivist Asian cultures (Bond, 2010). Our findings of

TABLE 4 | Correlations of the Family Communication Scale score with scores of 12-item Short Form Health Survey Version 2, Subjective Happiness Scale, family health, family harmony, family happiness, family APGAR scale, and perceived family communication quality and sufficiency in the population-based sample and community-based sample.

Correlation with the Family Communication Scale score (FCS; range 10–50)[a]	Population-based sample (*N* = 687)	Community-based sample (*N* = 1983)		
		Baseline (T1)	1-Month follow-up (T3)[b]	3-Month follow-up (T4)[b]
SF-12 Physical Component Subscale (PCS; range 0–100)	–	0.13	0.19	0.17
SF-12 Mental Component Subscale (MCS; range 0–100)	–	0.31	0.34	0.34
Subjective happiness scale (SHS; range 1–7)	0.43	0.47	0.48	0.46
Family health (range 0–10)	0.47	0.51	0.50	0.48
Family harmony (range 0–10)	0.56	0.61	0.59	0.58
Family happiness (range 0–10)	0.57	0.58	0.58	0.56
Family APGAR Scale (range 0–10)	0.49	–	–	–
Perceived family sufficiency (range 1–5)	0.40	–	–	–
Perceived family quality (range 0–10)	0.60	–	–	–

SF-12, 12-item Short Form Health Survey Version 2; APGAR, Adaption, Partnership, Growth, Affection, Resolve.
[a]All P-values for Pearson correlation coefficients < 0.001.
[b]Partial correlation was used to account for the intervention effect when assessing the correlations.

retaining all 10 items contrasted with this notion, suggesting a more direct exchange of information both factual and emotional within the Hong Kong Chinese population. Similar preference of explicit communication style was reported in our previous qualitative study (Lam et al., 2012). One possible explanation is that the Western media influences and the busy urban life have encouraged more direct and explicit communication in Hong Kong (Lam et al., 2012), the most westernized and modernized city in China.

Our results supported the FCS as a reliable and valid measurement of positive family communication in the Chinese population. The internal consistency reliability (Cronbach's alpha = 0.91; McDonald's Omega = 0.91) was good and comparable to those obtained in the original scale and other validations (Cronbach's alpha 0.88–0.92) (Olson and Barnes, 2004; Smith et al., 2009; Martínez-Pampliega et al., 2017; Rivadeneira and López, 2017). The 1-month test–retest reliability (intraclass correlation coefficient = 0.69) was moderate, despite the potential effects of tea gathering sessions on family communication in the control group. The correlation of positive family communication with mental HRQoL was significantly stronger than that with physical HRQoL. Although not directly measuring family communication, our previous study showed a stronger correlation of family happiness with mental HRQoL than physical HRQoL (Shen et al., 2019). Positive family communication was found to be positively and moderately correlated with individual happiness measured by SHS (Lyubomirsky and Lepper, 1999), family functioning measured by family APGAR (Smilkstein, 1978), and family well-being including health, harmony, and happiness developed specifically in the Chinese culture (Chan et al., 2011; Lam et al., 2012). The correlation of positive family communication with perceived communication quality was statistically stronger than that with perceived sufficiency. The difference implied that quality enhancement might be more important than the increase in family time to develop positive communication skills, particularly in Hong Kong where long work hours challenge shared family time (Wharton and Blair-Loy, 2006; Ho et al., 2018). A study among romantic partners also supported communication quality indicators (e.g., depth, smooth, social) but not quantity on predicting intimacy and relational satisfaction (Emmers-Sommer, 2004).

A higher household income was associated with greater positive family communication. High-income families tend to experience fewer financial problems particularly monetary difficulties that could induce stress and family conflicts (Orthner et al., 2004). Alternatively, people with high income were more likely to seek or share knowledge and skills to enhance family communication because of more social support, cognitive skills, and information literacy, which are documented barriers for people with low socioeconomic status (Wang et al., 2014; Shen et al., 2017b).

Frequent use of ICTs including phone calls, SNS, and IM with family members was associated with greater positive family communication. The perpetual connectivity pattern represented by ICTs can facilitate family communication in real time and/or be conducted asynchronously. For example, family members could take advantage of the mobility and immediacy to coordinate family activities through mobile devices during time on waiting or on the move (Lanigan, 2009). The pattern of media multiplexity by ICTs allows both factual and emotional information to be exchanged in a diversity of media such as texts, pictures, audio clips, and videos (Carvalho et al., 2015). Communication through phone calls could further provide instant feedback and multiple cues such as tones and inflection. Studies also suggested the potential adverse effects of ICTs on family communication, as ICTs might reduce communication content and context compared with the traditional face-to-face method that conveys verbal, non-verbal, and tacit knowledge simultaneously (Carvalho et al., 2015). We found that frequent face-to-face communication was strongly associated with positive family communication, which was consistent with our previous

TABLE 5 | Associations of sociodemographic characteristics with Family Communication Scale score in the population-based sample (N = 687).

	n (%)[a] (N = 693)	Mean FCS score (SD)[a,b]	Crude β (95% CI)	Adjusted[c] β (95% CI)
Gender				
Male	309 (44.5)	37.7 (5.5)	0	0
Female	385 (55.5)	37.9 (6.8)	0.26 (− 0.99, 1.50)	−0.24 (− 1.58, 1.10)
Age, years				
18–24	66 (9.5)	37.3 (7.1)	0	0
25–44	233 (33.6)	36.7 (3.5)	−0.59 (− 2.77, 1.59)	−1.80 (− 4.51, 0.91)
45–64	270 (38.9)	38.4 (6.2)	1.13 (− 0.91, 3.17)	−0.19 (− 3.02, 2.64)
≥65	125 (18.0)	38.8 (9.3)	1.48 (− 0.52, 3.47)	0.29 (− 2.80, 3.39)
P for trend			0.007	0.096
Marital status				
Unmarried	172 (24.8)	36.8 (6.0)	0	0
Cohabitated/married	460 (66.4)	38.2 (5.9)	1.35 (− 0.22, 2.92)	0.25 (− 1.96, 2.45)
Divorced/separated/widowed	61 (8.8)	37.7 (7.7)	0.85 (− 1.22, 2.92)	−0.16 (− 2.94, 2.62)
Employment status				
Unemployment	39 (5.6)	34.4 (3.5)	0	0
In-paid employment	328 (47.3)	37.6 (5.0)	3.26 (0.43, 6.08)	2.55 (− 0.40, 5.50)
Retired	152 (22.0)	38.5 (8.2)	4.15 (1.32, 6.97)**	2.67 (− 0.66, 6.00)
Housekeeper	134 (19.3)	38.8 (6.1)	4.43 (1.45, 7.41)**	3.57 (0.23, 6.92)*
Full-time student	41 (5.9)	36.8 (8.2)	2.47 (− 1.27, 6.21)	0.76 (− 3.37, 4.90)
Educational attainment				
Primary or below	171 (24.6)	38.6 (6.5)	0	0
Secondary	345 (49.7)	37.2 (5.3)	−1.38 (− 2.94, 0.17)	−0.51 (− 2.16, 1.14)
Tertiary	178 (25.6)	38.3 (7.4)	−0.32 (− 2.01, 1.36)	0.36 (− 1.71, 2.42)
P for trend			0.748	0.863
Monthly household income (HK $)[d]				
≤19,999	237 (34.2)	37.0 (6.2)	0	0
20,000–29,999	131 (18.8)	37.5 (6.3)	0.42 (− 1.65, 2.48)	0.42 (− 1.57, 2.41)
30,000–39,999	92 (13.3)	37.9 (6.1)	0.90 (− 1.08, 2.89)	1.37 (− 0.55, 3.29)
≥40,000	176 (25.4)	38.4 (5.3)	1.38 (− 0.03, 2.79)	1.42 (− 0.21, 3.05)
P for trend			0.052	0.045
Unstable/refused to answer	58 (8.3)	39.5 (7.7)	2.41 (0.54, 4.27)*	1.86 (− 0.14, 3.85)

FCS, Family Communication Scale, range 10–50.
$^{**}P < 0.01$; $^{*}P < 0.05$.
[a]Weighted by age, gender, and educational attainment distribution of the Hong Kong general population.
[b]Mean FCS score (SD) = 37.8 (6.2).
[c]Mutually adjusted for other variables in the table.
[d]US$ 1 = HK$ 7.8.

findings of the central role of face-to-face communication in improving family well-being (Wang et al., 2015; Shen et al., 2017b). Among the respondents who frequently conducted face-to-face family communication, frequent use of phone calls, and SNS were associated with even greater levels of positive family communication. Such findings supported that ICTs could be utilized as a supplement and extension to the traditional face-to-face communication method.

One of the limitations of Study 1 was the cross-sectional study design, which was subjected to residual confounding and restricted the inference of temporal sequence of the observed associations. We used the landline telephone sampling method, which excluded the increasing mobile phone-only households. The effects of non-response bias and coverage bias on the observed associations are uncertain. However, data were weighted according to gender, age, and educational attainment distribution of Hong Kong general population to increase the representativeness. We did not assess the geographical distance between family members, which could influence the selection of the communication method and frequency of use (Carvalho et al., 2015). All scales used in Study 1 and Study 2 were self-reported, which could be subjected to bias. Samples were from the Hong Kong Chinese population who have been exposed to social modernization, urban living, and Western cultural influences. Generalizability to rural settings and Chinese communities outside Hong Kong needs to be further studied. For example, measurement invariance tests of FCS can be used to assess the differences between urban and rural settings. Content validity and responsiveness of FCS were not evaluated. However, FCS scores increased with sustainable small effects up to 12 weeks in our previous interventions for enhancing family communication and well-being (Ho et al., 2016a,b).

TABLE 6 | Associations of communication method with Family Communication Scale score in the population-based sample (*N* = 687).

	n **(%)**	**Mean FCS score (SD)[a,b]**	**Crude β (95% CI)**	**Adjusted[c] β (95% CI)**	**Adjusted[c] β (95% CI) (*n* = 628)[d]**
Face to face					
Never/seldom	55 (8.0)	32.1 (8.4)	0	0	–
Sometimes/often	628 (92.0)	38.3 (5.7)	6.23 (3.57, 8.89)***	6.33 (3.71, 8.96)***	–
Phone call					
Never/seldom	227 (33.3)	36.0 (7.0)	0	0	0
Sometimes/often	455 (66.7)	38.7 (5.6)	2.65 (1.34, 3.96)***	2.74 (1.50, 3.97)***	1.99 (0.82, 3.15)***
Instant messaging (e.g., WhatsApp)					
Never/seldom	277 (40.6)	37.0 (7.0)	0	0	0
Sometimes/often	405 (59.4)	38.4 (5.5)	1.38 (0.09, 2.66)*	1.91 (0.53, 3.29)**	1.18 (-0.25, 2.61)
Video call (e.g., Skype, FaceTime, WeChat video call)					
Never/seldom	555 (81.4)	37.6 (6.5)	0	0	0
Sometimes/often	127 (18.6)	38.6 (4.6)	0.92 (−0.49, 2.34)	1.10 (−0.26, 2.45)	0.49 (−0.87, 1.86)
Social networking sites (e.g., Facebook)					
Never/seldom	595 (87.2)	37.5 (6.3)	0	0	0
Sometimes/often	87 (12.8)	39.5 (5.3)	1.95 (0.34, 3.57)*	2.26 (0.55, 3.96)**	1.84 (0.13, 3.57)*
Email					
Never/seldom	626 (91.8)	37.7 (6.2)	0	0	0
Sometimes/often	56 (8.2)	39.1 (5.5)	1.40 (−0.19, 2.99)	1.10 (−0.51, 2.72)	0.63 (−1.00, 2.26)

FCS, Family Communication Scale, range 10–50.
****P < 0.001; **P < 0.01; *P < 0.05.*
[a]Weighted by age, gender, and educational attainment distribution of the Hong Kong general population. [b]Mean FCS score (SD) = 37.8 (6.2).
[c]Adjusted for gender, age, marital status, employment status, educational attainment, and monthly household income.
[d]Among respondents who frequently conducted face-to-face family communication.

Our study suggests several avenues for future research. The stronger correlation of positive family communication with perceived family communication quality than sufficiency warranted qualitative research on content and context of communication among family members to provide a deeper understanding of the important role of communication quality. Longitudinal studies are needed to distill the causal relations between communication quality and positive family communication. Income inequalities in positive family communication warranted intervention studies for enhancing family communication specifically in low-income families. Our study is the first to show the ability of FCS to distinguish people having frequent family communication using ICTs, and ICTs could enhance family communication as supplements and extension of the traditional face-to-face method. The findings inform future digital health research in the family context that FCS can be an appropriate outcome measure.

CONCLUSION

We identified the one-factor structure of the Chinese version of FCS, which can serve as a valid and reliable measurement of positive family communication in the Chinese population. A higher monthly household income and frequent use of face-to-face communication and ICTs including phone calls, IM, and SNS were associated with greater positive family communication. Frequent use of phone calls and SNS can improve positive family communication among people who frequently conducted face-to-face family communication. The findings indicated that ICTs could be utilized as a supplement for traditional face to face to enhance family communication in the Chinese population.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the data that support the findings of this study are available from the FAMILY project but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the Hong Kong Jockey Club Charities Trust. Requests to access the datasets should be directed to jcfamily@hku.hk.

ETHICS STATEMENT

Ethical approvals of Study 1 and Study 2 were granted by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster. Verbal informed consent was obtained from the respondents in Study 1. Written consent was obtained from the adult participants in Study 2. For participants younger than 18 years in Study 2, written consent was obtained from the next of kin, caretakers, or guardians on their behalf.

AUTHOR CONTRIBUTIONS

MW, KV, SC, and THL conceived the study. NG analyzed the data from Study 1 and wrote the first draft of the manuscript. HH analyzed the data from Study 2. NG, HH, MW, and TTL interpreted the data. All authors critically revised and approved the final version of the manuscript.

FUNDING

This work was supported by the Hong Kong Jockey Club Charities Trust as part of the project: "FAMILY: a Jockey Club Initiative for a Harmonious Society" (https://www.family.org.hk/en). The publication fee was supported by Sir Robert Kotewall Endowed Professorship in Public Health Fund to THL.

ACKNOWLEDGMENTS

We thank the respondents in the telephone surveys and Public Opinion Program (HKU) for conducting the surveys. We thank our community collaborators and participants for their coordination, implementation, and participation in the community programs. We are grateful to THL, Sir Robert Kotewall Professorship in Public Health, for generous support for the publication fee.

REFERENCES

Aesaert, K., van Nijlen, D., Vanderlinde, R., and van Braak, J. (2014). Direct measures of digital information processing and communication skills in primary education: using item response theory for the development and validation of an ICT competence scale. *Comput. Educ.* 76, 168–181. doi: 10.1016/j.compedu.2014.03.013

Baiocco, R., Cacioppo, M., Laghi, F., and Tafà, M. (2013). Factorial and construct validity of FACES IV among Italian adolescents. *J. Child Fam. Stud.* 22, 962–970. doi: 10.1007/s10826-012-9658-1

Barnes, H. L., and Olson, D. H. (1985). Parent-adolescent communication and the circumplex model. *Child Dev.* 56, 438–447. doi: 10.2307/1129732

Bond, M. H. (ed.) (2010). *Oxford Handbook of Chinese Psychology*, 1st Edn. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199541850.001.0001

Carvalho, J., Francisco, R., and Relvas, A. P. (2015). Family functioning and information and communication technologies: how do they relate? A literature review. *Comput. Hum. Behav.* 45, 99–108. doi: 10.1016/j.chb.2014.11.037

Challier, B., Chau, N., Prédine, R., Choquet, M., and Legras, B. (2000). Associations of family environment and individual factors with tobacco, alcohol and illicit drug use in adolescents. *Eur. J. Epidemiol.* 16, 33–42.

Chan, D. H., Ho, S. C., and Donnan, S. P. B. (1988). A survey of family APGAR in Shatin private ownership homes. *Hong Kong Pract.* 10, 3295–3299.

Chan, S. S., Viswanath, K., Au, D. W. H., Ma, C. M. S., Lam, W. W. T., Fielding, R., et al. (2011). Hong Kong Chinese community leaders' perspectives on family health, happiness and harmony: a qualitative study. *Health Educ. Res.* 26, 664–674. doi: 10.1093/her/cyr026

Comrey, A. L., and Lee, H. B. (2013). *A First Course in Factor Analysis*. Hove: Psychology press.

Courtney, M. G. R., and Gordon, M. (2013). Determining the number of factors to retain in EFA: using the SPSS R-Menu v2. 0 to make more judicious estimations. *Pract. Assess. Res. Eval.* 18, 1–14.

Dinno, A. (2009). Implementing Horn's parallel analysis for principal component analysis and factor analysis. *Stata J.* 9, 291–298.

Emmers-Sommer, T. M. (2004). The effect of communication quality and quantity indicators on intimacy and relational satisfaction. *J. Soc. Pers. Relat.* 21, 399–411. doi: 10.1177/0265407504042839

Gomes, H. M. S., Peixoto, F., and Gouveia-Pereira, M. (2017). Portuguese validation of the family adaptability and cohesion evaluation scale–FACES IV. *J. Fam. Stud.* 25, 1–18. doi: 10.1080/13229400.2017.1386121

Gonzalez, C., and Katz, V. S. (2016). Transnational family communication as a driver of technology adoption. *Int. J. Commun.* 10:21.

Hayes, A. F., and Coutts, J. J. (2020). Use Omega rather than Cronbach's Alpha for estimating reliability. *Commun. Methods Meas.* 14, 1–24. doi: 10.1080/19312458.2020.1718629

Ho, H. C. Y., Mui, M., Wan, A., Ng, Y., Stewart, S. M., Yew, C., et al. (2016a). Happy family kitchen: a community-based research for enhancing family communication and well-being in Hong Kong. *J. Fam. Psychol.* 30, 752–762. doi: 10.1037/fam0000233

Ho, H. C. Y., Mui, M., Wan, A., Ng, Y.-L., Stewart, S. M., Yew, C., et al. (2016b). Happy family kitchen II: a cluster randomized controlled trial of a community-based family intervention for enhancing family communication and well-being in Hong Kong. *Front. Psychol.* 7:638. doi: 10.3389/fpsyg.2016.00638

Ho, H. C. Y., Mui, M., Wan, A., Stewart, S. M., Yew, C., Lam, T. H., et al. (2017). Happy family kitchen: behavioral outcomes of a brief community-based family intervention in Hong Kong. *J. Child Fam. Stud.* 26, 2852–2864. doi: 10.1007/s10826-017-0788-3

Ho, H. C. Y., Mui, M., Wan, A., Yew, C., Lam, T. H., Chan, S. S., et al. (2018). Family meal practices and well-being in Hong Kong: the mediating effect of family communication. *J. Fam. Issues* 39, 3835–3856. doi: 10.1177/0192513X18800787

Ho, H. C. Y., Mui, M. W., Wan, A., Yew, C. W., and Lam, T. H. (2019). Happy family kitchen movement: a cluster randomized controlled trial of a community-based family holistic health intervention in Hong Kong. *J. Happiness Stud.* 21, 15–36. doi: 10.1007/s10902-018-00071-w

Hooper, D., Coughlan, J., and Mullen, M. (2008). Structural equation modelling: guidelines for determining model fit. *Electron. J. Bus. Res.* 6:9.

Izrael, D., Hoaglin, D. C., and Battaglia, M. P. (2004). "To rake or not to rake is not the question anymore with the enhanced raking macro," in *Proceedings of the 29th Annual SAS Users Group International Conference*, Montreal, CA.

Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor. Dent. Endod.* 38, 52–54. doi: 10.5395/rde.2013.38.1.52

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling*, 4th Edn. New York, NY: Guilford Publications.

Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012

Koutra, K., Triliva, S., Roumeliotaki, T., Lionis, C., and Vgontzas, A. N. (2013). Cross-cultural adaptation and validation of the greek version of the family adaptability and cohesion evaluation scales iv package (FACES IV Package). *J. Fam. Issues* 34, 1647–1672. doi: 10.1177/0192513X12462818

Lam, C. L. K., Tse, E. Y. Y., and Gandek, B. (2005). Is the standard SF-12 health survey valid and equivalent for a Chinese Population? *Qual. Life Res.* 14, 539–547.

Lam, W. W. T., Fielding, R., McDowell, I., Johnston, J., Chan, S., Leung, G. M., et al. (2012). Perspectives on family health, happiness and harmony (3H) among Hong Kong Chinese people: a qualitative study. *Health Educ. Res.* 27, 767–779. doi: 10.1093/her/cys087

Lanigan, J. D. (2009). A sociotechnological model for family research and intervention: how information and communication technologies affect family life. *Marriage Fam. Rev.* 45, 587–609. doi: 10.1080/01494920903224194

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* 48, 936–949. doi: 10.3758/s13428-015-0619-7

Lyubomirsky, S., and Lepper, H. S. (1999). A measure of subjective happiness: preliminary reliability and construct validation. *Soc. Indic. Res.* 46, 137–155.

Martínez-Pampliega, A., Merino, L., and Iriarte, L. (2017). Psychometric properties of the Spanish version of the family adaptability and cohesion evaluation scale IV. *Psicothema* 29, 414–420. doi: 10.7334/psicothema2016.21

Moret-Tatay, C., Beneyto-Arrojo, M. J., Gutierrez, E., Boot, W. R., and Charness, N. (2019). A Spanish adaptation of the computer and mobile device proficiency questionnaires (CPQ and MDPQ) for Older Adults. *Front. Psychol.* 10:1165. doi: 10.3389/fpsyg.2019.01165

Nan, H., Ni, M. Y., Lee, P. H., Tam, W. W. S., Lam, T. H., Leung, G. M., et al. (2014). Psychometric evaluation of the Chinese version of the subjective happiness scale: evidence from the Hong Kong FAMILY Cohort. *Int. J. Behav. Med.* 21, 646–652. doi: 10.1007/s12529-014-9389-3

Olson, D. H. (2000). Circumplex model of marital and family sytems. *J. Fam. Ther.* 22, 144–167. doi: 10.1111/1467-6427.00144

Olson, D. H., and Barnes, H. L. (2004). "Family communication," in *FACES IV Package* (Minneapolis, MN: Life Innovations).

Orthner, D. K., Jones-Sanpei, H., and Williamson, S. (2004). The resilience and strengths of low-income families. *Fam. Relat.* 53, 159–167.

Rivadeneira, J., and López, M. A. (2017). Family communication scale: validation in Chilean. *Acta Colomb. Psicol.* 20, 127–137.

Schneider, L. A., King, D. L., and Delfabbro, P. H. (2017). Family factors in adolescent problematic Internet gaming: a systematic review. *J. Behav. Addict.* 6, 321–333. doi: 10.1556/2006.6.2017.035

Schrodt, P., Witt, P. L., and Messersmith, A. S. (2008). A meta-analytical review of family communication patterns and their associations with information processing, behavioral, and psychosocial outcomes. *Commun. Monogr.* 75, 248–269. doi: 10.1080/03637750802256318

Shen, C., Wan, A., Kwok, L. T., Pang, S., Wang, X., Stewart, S. M., et al. (2017a). A community-based intervention program to enhance family communication and family well-being: the learning families project in Hong Kong. *Front. Public Health* 5:257. doi: 10.3389/fpubh.2017.00257

Shen, C., Wang, M. P., Chu, J. T., Wan, A., Viswanath, K., Chan, S. S. C., et al. (2017b). Sharing family life information through video calls and other information and communication technologies and the association with family well-being: population-based survey. *JMIR Ment. Health* 4:e57. doi: 10.2196/mental.8139

Shen, C., Wang, M. P., Ho, H. C. Y., Wan, A., Stewart, S. M., Viswanath, K., et al. (2019). Test-retest reliability and validity of a single-item self-reported family happiness scale in Hong Kong Chinese: findings from Hong Kong Jockey Club FAMILY Project. *Qual. Life Res.* 28, 535–543. doi: 10.1007/s11136-018-2019-9

Smilkstein, G. (1978). The family APGAR: a proposal for a family function test and its use by physicians. *J. Fam. Pract.* 6, 1231–1239.

Smith, K. M., Freeman, P. A., and Zabriskie, R. B. (2009). An examination of family communication within the core and balance model of family leisure functioning. *Fam. Relat.* 58, 79–90. doi: 10.1111/j.1741-3729.2008.00536.x

Tabachnick, B. G., Fidell, L. S., and Ullman, J. B. (2007). *Using Multivariate Statistics*. Boston, MA: Pearson.

Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *J. Diagn. Med. Sonogr.* 6, 35–39.

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *J. Clin. Epidemiol.* 60, 34–42. doi: 10.1016/j.jclinepi.2006.03.012

Türkdoğan, T., Duru, E., and Balkıs, M. (2018). Turkish adaptation of the family adaptability and cohesion scale IV. *Int. J. Assess. Tools Educ.* 5, 631–644. doi: 10.21449/ijate.409110

Velicer, W. F., Eaton, C. A., and Fava, J. L. (2000). "Construct explication through factor or component analysis: a review and evaluation of alternative procedures for determining the number of factors or components," in *Problems and Solutions in Human Assessment*, eds R. D. Goffin and E. Helmes (Boston, MA: Springer), 41–71.

Wang, M. P., Chu, J. T., Viswanath, K., Wan, A., Lam, T. H., and Chan, S. S. (2015). Using information and communication technologies for family communication and its association with family well-being in Hong Kong: FAMILY project. *J. Med. Internet Res.* 17:e207. doi: 10.2196/jmir.4722

Wang, M. P., Wang, X., Viswanath, K., Wan, A., Lam, T. H., and Chan, S. S. (2014). Digital inequalities of family life information seeking and family well-being among chinese adults in Hong Kong: a population survey. *J. Med. Internet Res.* 16:e227. doi: 10.2196/jmir.3386

Ware, J., Kosinski, M., and Keller, S. D. (1996). A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med. Care* 34, 220–233.

Wharton, A. S., and Blair-Loy, M. (2006). Long work hours and family life: a cross-national study of employees' Concerns. *J. Fam. Issues* 27, 415–436. doi: 10.1177/0192513X05282985

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 November 2021
doi: 10.3389/fpsyg.2021.774472

Check for updates

Review of the Internal Structure, Psychometric Properties, and Measurement Invariance of the Work-Related Rumination Scale – Spanish Version

Ernesto Rosario-Hernández[1,2], Lillian V. Rovira-Millán[3] and César Merino-Soto[4]*

[1] Clinical Psychology Programs, School of Behavioral and Brain Sciences, Ponce Health Sciences University, Ponce, PR, United States, [2] Ponce Research Institute, Ponce Health Sciences University, Ponce, PR, United States, [3] Psychology Program, Department of Social Sciences, University of Puerto Rico, Cayey, PR, United States, [4] Psychology Research Institute, School of Psychology, University of San Martín de Porres, Lima, Peru

OPEN ACCESS

Edited by:
Victor Zaia,
Faculdade de Medicina do ABC, Brazil

Reviewed by:
Hamid Sharif Nia,
Mazandaran University of Medical Sciences, Iran
Maria Do Carmo Fernandes Mratins,
Universidade Metodista de São Paulo (UMESP), Brazil

****Correspondence:***
Ernesto Rosario-Hernández
erosario@psm.edu

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *12 September 2021*
Accepted: *19 October 2021*
Published: *25 November 2021*

Citation:
Rosario-Hernández E, Rovira-Millán LV and Merino-Soto C (2021) Review of the Internal Structure, Psychometric Properties, and Measurement Invariance of the Work-Related Rumination Scale – Spanish Version. Front. Psychol. 12:774472. doi: 10.3389/fpsyg.2021.774472

Background: The aim of the current study was to examine the internal structure and assess the psychometric properties of the Work-Related Rumination Scale (WRRS) – Spanish version in a Puerto Rican sample of workers. This instrument is a 15-item questionnaire, which has three factors, affective rumination, problem-solving pondering, and detachment. This measure is used in the occupational health psychology context; however, there is little evidence of its psychometric properties.

Materials and Methods: A total sample of 4,100 from five different study samples was used in this cross-sectional study design in which the WRRS was used. We conducted confirmatory factor analysis (CFA) and exploratory structural equation modeling (ESEM) to examine the internal structure of the Work-Related Rumination Scale. Measurement invariance across sex and age was examined.

Results: The three-factor model was supported; however, four items were eliminated due to their cross-loadings and factorial complexity. This 11-item Spanish version of the WRRS was invariant across sex and age. Reliability of the three-factors of WRRS were within the range of 0.74 to 0.87 using Cronbach's alpha and McDonald's omega. Correlations between the three factors were as expected as well as with other established measures.

Conclusion: The results suggest that the WRRS-Spanish version appears to be a reliable and valid instrument to measure work-related rumination using its three factors. Comparison across sex and age appear to be useful in occupational health psychology research setting since results suggest that the WRRS is invariant regarding those variables.

Keywords: rumination, ESEM, CFA, invariance, detachment, problem-solving pondering

INTRODUCTION

The link between the exposure to work demands and the possible deterioration of employee's health is an area of interest for occupational stress research (Pereira and Elfering, 2014). Work demands have been associated to a series of health complications such as cardiovascular diseases (Karasek et al., 1981; Rosario-Hernández et al., 2014), burnout (Brotheridge and Grandey, 2002), depression (Dormann and Zapf, 1999; Blackmore et al., 2007; Magnavita and Fileni, 2014; Rosario-Hernández et al., 2014), and psychosomatic symptoms (Pisanti et al., 2003; van der Doef et al., 2012; Rosario-Hernández et al., 2013).

On the other hand, impediments to recovering from work demands can impair employee's health (Meijman and Mulder, 1998; Schwartz et al., 2003; Kivimäki et al., 2006; Zijlstra and Sonnentag, 2006; Fritz et al., 2010). Thus, the process of recovery appears to be influenced in the way in which people can disconnect from their work demands and those thoughts related to them (Cropley et al., 2006; Rook and Zijlstra, 2006; Sonnentag and Zijlstra, 2006; Sonnentag et al., 2008). In this way, recovery from work is necessary for workers to avoid chronic stress (Safstrom and Harting, 2013) and therefore, rumination is a mechanism suggested that can compromise a successful disconnection and recovery from work (Roger and Jamieson, 1988; Cropley et al., 2006). Cropley and Zijlstra (2011) indicate that work-related rumination can be considered as a set of repetitive thoughts directed to issues that revolve around work; it does not matter, really, if people ruminate or think about work issues when not at work and in fact, many people do it because find it rewarding and stimulating. However, Cropley and Zijlstra (2011) argue that rumination becomes a problem when affects health and well-being. Thus, Cropley and Zijlstra suggest that people not always worry or think negatively about work on their off time. In fact, thinking about work is not compatible to switch off, and therefore, makes it difficult to recover from work. On the other hand, thinking and reflexing about work issues can also have beneficial effects and can be associated to positive results.

Furthermore, Cropley and Zijlstra (2011) conceptualize work-related rumination as a construct with three factors, which they call affective rumination (AR), problem-solving pondering (PSP), and detachment (Det). AR is a cognitive state characterize by the appearance intrusive, penetrating, and recurrence thoughts about work. These thoughts are negative in affective terms (Pravettoni et al., 2007), which if are not controlled, can become cognitively and emotionally intrusive thoughts when off work. Meanwhile, Cropley and Zijlstra point out that most of studies related to rumination at work have focused on its negative aspect, which imply if people continue to think about their work when off, they continue to be with the "power button on" and this prevent them to recuperate during their off time. It is very clear that this type of rumination impact negatively recovery when not at work; however, thinking about work when not on it, not necessarily have negative implications, since it may have a positive side. For example, there are studies that suggest that thinking about work when off might have a positive impact on innovation and creativity (e.g., Baas et al., 2008). For instance, the results obtained by Baas et al. suggest that people tend to have a positive humor when the task at hand was found to be pleasantly and intrinsically helpful. Similarly, PSP, according to Cropley and Zijlstra (2011), is a mode of thinking characterized by lengthy mental examination or the appraisal of a past difficulty at work in order to discover a solution. Finally, detachment is the third factor of the work-related rumination, and it can be defined as a sense of being away from the work situation (e.g., Etzion et al., 1998). Cropley and Zijlstra (2011) indicate that there are people who manage to press the "off button" and can disconnect and forget about work.

Based on this conceptualization, Cropley et al. (2012) developed the Work-Related Rumination Scale (WRRS), which has been used in occupational health psychology research and has been translated into different languages to measure rumination at work in different studies. These translations have been done by Syrek et al. (2017) into German, Firoozabadi et al. (2018a) into Persian, Sulak Akyüz and Sulak (2019) into Turkish, and in Puerto Rico by Rosario-Hernández et al. (2013) into Spanish. The confirmatory factor analysis (CFA) results obtained of these translations of the WRRS are like those obtained on research by Cropley et al. (2012) and Querstret and Cropley (2012) because they also yielded a three-factor internal structure; AR, PSP, and detachment.

Brief Systematic Literature Review of the Work-Related Rumination Scale

A brief systematic review was conducted to establish the pattern of findings and methodological procedures used in studies of the psychometric properties in general, and internal structure of the WRRS, as recommended by some authors in the literature (e.g., Grant and Booth, 2009). The following key words were used: WRRS AND internal structure OR psychometric properties AND validity AND reliability OR measurement invariance The review was done through the search engines in the EBSCO, Sciencedirect, Scopus, Pubmed, and Google Scholar databases, using "Boolean" connectors between November 2020 and May 2021. Our intention, at first, was to include only studies about psychometric properties of the WRRS, but given that we only found one with at least some variety of validity evidence, it was decided to include studies which at least tested for some sort of psychometric property as part of the study, such as those that used structural equation modeling (SEM) as an analytical tool in which was tested the measurement model and those who at least reported the reliability of the WRRS (see **Table 1**). Thus, we only found one study in which its main research objective was to examine the psychometric properties of WRRS (Sulak Akyüz and Sulak, 2019). This mentioned study was the Turkish version of the WRRS, and their CFA results supported the three-factor model proposed by the WRRS's authors using the maximum likelihood estimation. Also, they reported reliability coefficients ranging between 0.73 to 0.79 and appears that they did not examine for measurement invariance because it was not reported.

Interestingly, of the 25 studies revised, only seven studies used the complete WRRS and those included the original study in which the WRRS was developed (Cropley et al., 2012; Querstret and Cropley, 2012; Vandevala et al., 2017;

TABLE 1 | Brief literature review of the work-related rumination scale.

Study-Country-Version	Study Main Objective	Participants	Factorial Design	Factorial Loading	Method	Factor Relationship	Internal Consistency	Invariance
(1) Cropley et al. (2012) United Kingdom Original Scale (English)	Not psychometric	$n = 268$ **Gender**: Females –58.6% Males-41.4% **Age:** 19–63 $M = 36.7$ $SD = 12.9$	EFA	**3 factors**: AR PSP Det	**Estimator**: NR **Rotation**: Direct Oblimin	AR←→PSP: 0.61 AR←→Det: –0.63 PSP←→Det: –0.51	**Cronbach's Alpha**: AR = 0.90 PSP = 0.82 Det = 0.86	NR
(2) Querstret and Cropley (2012) United Kingdom Original Scale (English)	Not psychometric	$n = 719$ **Gender**: Females (49.2%) **Age:** 19–69 Female: $M = 32$, $SD = 10.5$ Males: $M = 35$, $SD = 10.7$	NR	NR	NR	**Zero Order Correlations:** NR	**Cronbach's Alpha**: PSP = 0.80 Det = 0.83	NR
(3) Zoupanou et al. (2013)	Not psychometric	$n = 310$ **Gender**: Females (50%) Males (50% **Age:** 19–69 $M = 42.91$ $SD = 9.41$	NR	NR	NR	**Zero Order Correlations:** AR←→PSP:0.32	**Cronbach's Alpha**: PSP = 0.80 Det = 0.83	NR
(4) Querstret et al. (2016) United Kingdom Original Scale (English)	Not psychometric	$n = 227$ **Gender**: Females (63.0%) Males (37.0%) **Age:** 22–66 $M = 42.62$ $SD = 9.83$	NR	NR	NR	NR	**Cronbach's Alpha**: AR $T_1 = 0.85$ $T_2 = 0.87$	NR
(5) Syrek et al. (2017) Germany German Version	Not psychometric	$n = 357$ **Gender**: Females (76.0%) **Age:** 21–59 $M = 36.0$ $SD = 9.4$	CFA	**2 factors**: AR PSP	**Estimator**: NR **Model**: 1 and 2 Factors	AR←→PSP:0.64	**Cronbach's Alpha**: AR = 0.91 PSP = 0.84	NR
(6) Vahle-Hinz et al. (2017) Germany and Finland German Version	Not psychometric	$n_{T1} = 1{,}347$ $n_{T2} = 841$ $n_{T3} = 630$ **Gender:** n_{T1} = Females (58.0%) **Age:** 23–66 n_{T1}: $M = 47.5$ $SD = 9.9$	CFA	**2 Factors:** AR PSP	**Estimator**: MLR **Model**: AR PSP	**Zero Order Correlations:** T_1: AR←→PSP:0.19 T_2: AR←→PSP:0.13 T_3: AR←→PSP:0.07	**Cronbach's Alpha**: T_1: AR = 0.87 T_1: PSP = 71 T_2: AR = 0.89 T_2: PSP = 0.74 T_3: AR = 0.89 T_3: PSP = 0.70	NR
(7) Bisht (2017) India Version NR	Not psychometric	$n = 297$ **Gender:** NR **Age:** 20–35	CFA	**2 Factors:** AR PSP	**Estimator**: NR **Model**: 2 factors	**Zero Order Correlations:** AR←→PSP:0.45	**Cronbach's Alpha**: AR = 0.87 PSP = 0.83	NR
(8) Vandevala et al. (2017) United Kingdom Original Version	Not psychometric	$n = 96$ **Gender:** Females – 52% Males – 47.9% **Age:** 31–50 years	NR	NR	NR	**Zero Order Correlations:** AR←→PSP:0.45 AR←→Det:0.50 PSP←→Det:0.35	**Cronbach's Alpha**: AR = 0.83 PSP = 0.43 Det = 0.76	NR
(9) Querstret et al. (2017) United Kingdom Original Version	Not psychometric	$n = 118$ **Gender:** Females – 80.5% **Age:** 21–62 $M = 40.68$ $SD = 10.45$	NR	NR	NR	**Zero Order Correlations:** NR	**Cronbach's Alpha**: AR = 0.85/0.87/0.89/0.89 PSP = 0.70/0.74/0.71/0.78	NR
(10) Svetieva et al. (2017) United States Original Version	Not psychometric	$n = 384$ **Gender:** Females – 48% **Age:** 35–65	NR	NR	NR	NR	**Cronbach's Alpha**: Det = 0.73	NR
(11) Kinnunen et al., 2017	Not psychometric	$n = 841$ **Gender:** Females – 58.6% **Age:** 21–67 $M = 47.1$, $SD = 10.00$	CFA	**2 Factors:** AR PSP	**Estimator**: MLR **Model**: AR PSP	**Zero Order Correlations:** T_1:AR←→PSP:0.18 T_2:AR←→PSP:0.15	**Cronbach's Alpha**: AR = 0.88/0.89 PSP = 0.70/0.69	NR

(Continued)

TABLE 1 | (Continued)

Study-Country-Version	Study Main Objective	Participants	Factorial Design	Factorial Loading	Method	Factor Relationship	Internal Consistency	Invariance
(12) Firoozabadi et al. (2018a) Iran Persian Version	Not psychometric	n = 123 **Gender**: Females –49.5% Males-50.4% **Age:** 21–54 M = 32.85, SD = 5.82	CFA	**2 factors**: AR PSP	**Estimator**: ML **Model**: 1 and 2 Factors	AR←→PSP:0.35	**Cronbach's Alpha**: AR = 0.87 PSP = 0.90	NR
(13) Firoozabadi et al. (2018b) Iran Persian Version	Not psychometric	n = 171 **Gender**: Females –55% Males-45% **Age:** 21–54 M = 32.75 SD = 5.67	CFA	**2 factors**: AR PSP	NR	**Zero Order Correlations:** AR←→PSP:0.40	**Cronbach's Alpha**: AR = 0.87 PSP = 0.90	NR
(14) Van Laethem et al. (2019) Finland		n = 920 **Gender**: Females –62.5% Males-45% **Age:** 40–60 M = 47.26 SD = 9.79	NR	NR	NR	NR	**Cronbach's Alpha**: AR = 0.87/0.89	NR
(15) Sulak Akyüz and Sulak (2019) Turkish Version	Psychometric	n = 582 **Gender**: Females –45.0% Males-55.0% **Age:** 21–59 M = 36.64 SD = 9.99	CFA	**3 factors**: AR: 1, 5, 7, 9, 15 PSP: 2, 4, 8, 11, 13 Det: 3, 6, 10, 12, 14	**Estimator**: ML **Model**: 3 Factors	AR←→PSP:0.66 AR←→Det: –0.69 PSP←→Det: –0.82	**Cronbach's Alpha**: AR = 0.79 PSP = 0.73 Det = 0.79	NR
(16) Weigelt et al. (2019a) Germany German Version	Not psychometric	n = 474 **Gender:** Females-61.8% **Age:** 20–59 M = 37.04 SD = 9.41	CFA	**5 Factors:** AR PSP Det PWR NWR	**Estimator:** DWLS Model: 1, 3, 4a, 4b, & 5	AR←→PSP:0.53 AR←→Det: –0.64 PSP←→Det: –0.62 AR←→PWR: –0.12 AR←→NWR:0.64 PSP←→PWR:0.44 PSP←→NWR:0.40 Det←→PWR: –0.02 Det←→NWR: –0.40	**Cronbach's Alpha**: AR = 0.90 PSP = 0.82 Det = 0.85	NR
(17) Weigelt et al. (2019b) Germany German Version	Not psychometric	n = 68 **Gender:** Females-72% Males-28% **Age:** 19–72 M = 34.34, SD = 9.55	CFA	NR	NR	NR	**Cronbach's Alpha**: AR = 0.84/0.95	NR
(18) Dunn and Sensky (2018) United Kingdom Original Version	Not psychometric	n = 79 **Gender:** Females-72% Males-28% **Age:** 19–72 M = 34.34 SD = 9.55	NR	NR	NR	**Zero Order Correlations:** AR←→PSP:0.41 AR←→Det: –0.58 PSP←→Det: –0.56	**Cronbach's Alpha**: AR = 0.91 PSP = 0.81 Det = 0.75	NR
(19) Kinnunen et al. (2019) Finland Finnish Version	Not psychometric	n_1 = 1,347 n_2 = 841 n_3 = 664 **Gender:** Females-58.0% **Age:** 23–66 M = 47.5 SD = 9.9	CFA	**Factors:** *AR* *PSP* Det	**Estimator:** MLR Model: 3 factors	**Zero Order Correlations:** T_1:AR←→PSP:0.19 T_1:AR←→Det: –0.33 T_1:PSP←→Det: –0.55 T_2:AR←→PSP:0.13 T_2:AR←→Det: –0.31 T_2:PSP←→Det: –0.49 T_2:AR←→PSP:0.19 T_3: AR←→PSP:0.07 T_3:AR←→Det: –0.29 T_3:PSP←→Det: –0.51	**Cronbach's Alpha**: AR = 0.87,0.89,0.89 PSP = 0.68,0.68.70	NR
(20) Cropley and Collis (2020) United Kingdom Original Scale (English)	Not psychometric	n = 104 **Gender:** Males (52.9%) **Age:** 19–66 M = 33.2 SD = 10.86	NR	NR	NR	NR	**Cronbach's Alpha**: AR = 0.87	NR

(Continued)

TABLE 1 | (Continued)

Study-Country-Version	Study Main Objective	Participants	Factorial Design	Factorial Loading	Method	Factor Relationship	Internal Consistency	Invariance
(21) Zhang et al. (2020) Germany German Version	Not psychometric	*n* = 1,109 **Gender:** Females (55.2%) Males (44.8%) **Age:** 18–65 *M* = 34.02 *SD* = 10.57	CFA	Factors: AR PSP	NR	**Zero Order Correlations:** AR←→PSP:0.31	**Cronbach's Alpha**: AR = 0.89 PSP = 0.86	NR
(22) Mullen et al. (2020) United States English Version (Original)	Not psychometric	*n* = 288 **Gender:** Females (81.9%) **Age:** 24–71 *M* = 43.37 *SD* = 11.19	NR	NR	NR	AR←→PSP:0.33 AR←→Det: –0.68 PSP←→Det: –0.40	**Cronbach's Alpha**: AR = 0.94 PSP = 0.92 Det = 0.96	NR
(23) Junker et al. (2020) Germany German Version	Not psychometric	*n* = 519 **Gender:** Females (90.6%) **Age:** *M* = 37.58 *SD* = 7.97	CFA	NR	NR	**Zero Order Correlations:** AR←→PSP:0.56	**Cronbach's Alpha**: AR = 0.85 PSP = 0.68 **McDonald's Omega:** AR = 0.85 PSP = 0.72	NR
(24) Pauli and Lang (2021) Germany German Version	Not psychometric	*n* = 1,836 **Gender:** NR **Age:** NR	NR	NR	NR	**Zero Order Correlations:** AR←→PSP:0.45	**Cronbach's Alpha**: AR = 0.90 PSP = 0.81	NR
(25) Mehmood and Hamstra (2021) Pakistan Original Version (English)	Not psychometric	*n* = 300 **Gender:** Females (90.6%) **Age:** 21–58 *M* = 30.76 *SD* = 7.42	CFA	NR	NR	NR	**Cronbach's Alpha**: PSP = 0.86	NR

NR, not reported; EFA, exploratory factor analysis; CFA, confirmatory factor analysis; ML, maximum likelihood; MLR, robust maximum likelihood; DWLS, diagonally weight least square; M, mean; SD, standard deviation; AF, affective rumination; PSP, problem-solving pondering; Det, Detachment.

Dunn and Sensky, 2018; Sulak Akyüz and Sulak, 2019; Weigelt et al., 2019a; Mullen et al., 2020), 11 used the affective and problem-solving pondering subscales (Bisht, 2017; Kinnunen et al., 2017, 2019; Querstret et al., 2017; Syrek et al., 2017; Vahle-Hinz et al., 2017; Firoozabadi et al., 2018a,b; Junker et al., 2020; Zhang et al., 2020; Pauli and Lang, 2021), two studies used the problem-solving pondering and detachment subscales (Zoupanou et al., 2013; Mehmood and Hamstra, 2021), only one used the detachment subscale (Svetieva et al., 2017), and four studies used the affective rumination subscale (Querstret et al., 2016; Van Laethem et al., 2019; Weigelt et al., 2019b; Cropley and Collis, 2020; Smyth et al., 2020). Thus, the use of the subscales of the WRRS vary according to the researchers need and purpose. But the use of affective rumination and problem-solving pondering are the most widely used subscales of the WRRS.

Regarding of method of factorial designs, one used exploratory factor analysis (EFA; Cropley et al., 2012), seven studies used CFA (Bisht, 2017; Syrek et al., 2017; Vahle-Hinz et al., 2017; Firoozabadi et al., 2018a; Kinnunen et al., 2019; Sulak Akyüz and Sulak, 2019; Weigelt et al., 2019a,b; and two of the studies did not report any of such methods, Querstret et al., 2016; Cropley and Collis, 2020). Those seven studies that relied on CFA, two studies used the maximum likelihood (ML) estimator, two used robust maximum likelihood (MLR), one used diagonally-weight least squares (DWLS), and two did not report it. Moreover, none of the studies examined the internal structure using exploratory structural equation modeling (ESEM) and none examined the measurement invariance of the WRRS. In addition, and in terms of the examination of the internal consistency, all the studies used Cronbach's alpha, and only one (Junker et al., 2020) used McDonald's omega that is a better estimate for the internal consistency (Crutzen and Peters, 2017; Flora, 2020).

Another point that stands out from the brief systematic review of the WRRS is that in the studies that did not use SEM, they presumed that the WRRS was a valid instrument without examining it with their sample. This tends to be a bad practice widely used in psychological studies, which has been pointed out by some authors (Merino-Soto and Calderón-De la Cruz, 2018; Merino-Soto and Angulo-Ramos, 2020, 2021) in the literature indicating that researchers are inducing the validity of the instrument, which is called as measurement validity induction.

Therefore, an attempt was made to push forward the research of the internal structure of the WRRS by implementing ESEM (Asparouhov and Muthén, 2009) approach, a model not incorporated in previous studies of the dimensionality of the WRRS, which is a reformulation of the modeling of item-construct relationships to solve CFA modeling problems. ESEM provides more information to decide on the multidimensionality of a measure created to represent multidimensional constructs (Morin et al., 2015). The ESEM was developed to subsume the exploratory approach within SEM, and characteristically consists of estimating the cross-factorial loads in the rest of the factors analyzed, and not only in the factor hypothesized as the main causal influence of the items (Asparouhov and Muthén, 2009).

The implementation of a traditional exploratory approach, as occurs in some studies with the WRRS does not seem to be different from the ESEM, because in the exploratory model's factor loadings are also estimated in all the factors. However, the advantages of nested exploratory modeling in SEM lead to obtaining fit measures, examining correlated residuals and other parameters usually not estimated in the exploratory approach (Asparouhov and Muthén, 2009; Mansolf and Reise, 2016). Estimates through ESEM have been shown to influence the decrease in factor loadings and interfactorial correlations (Asparouhov and Muthén, 2009; Mansolf and Reise, 2016). In this way, the factorial solutions obtained by the ESEM approach are considered more realistic (Asparouhov and Muthén, 2009). Due to the consistent demonstration of the efficacy of representing multidimensional constructs by means of the ESEM, the results of validation of the internal structure of the WRR in previous studies may present important biases in its parameters (i.e., factor loadings and latent correlations).

This assessment of WRR dimensionality even necessary when only estimating a reliability coefficient (specifically, internal consistency), for non-psychometric objective purposes, because proper estimation of reliability requires factor modeling (Crutzen and Peters, 2017; Flora, 2020). Studies that did not estimate reliability coefficients with their data generally induce reliability from other studies (Vassar et al., 2008), but there is no guarantee that the value obtained by inducing it from another study is equal to the one that could be calculated on their own data. On the other hand, equivalence of measurement between groups is required to ensure comparisons between groups with respect to statistics of interest, such as means, variances and covariation between scores.

In the same way, other aspects are also useful to examine for the quality of the instrument, such as the consistency of individual response (e.g., the items), especially when it is required to select items for the construction or adaptation of measures (Zijlmans et al., 2019), and that they are estimated within a reliability framework at the item level. Reliability is commonly estimated for the composite scores of the dimension constructed by the items; however, the reliability of the items is relevant for knowing the degree of reproducibility of the responses and has recently been valued as a quality measure for the choice of the items (Zijlmans et al., 2019).

Research Purpose

The WRRS was translated into Spanish and has been used in several studies in occupational health psychology in Puerto Rico (Rosario-Hernández et al., 2013, 2015, 2018a,b, 2019, 2020); however, psychometric properties of the WRRS have not been examined. Therefore, the purpose of the current study was to examine the internal structure, psychometric properties, and measurement invariance of the WRRS – Spanish version across gender and age.

MATERIALS AND METHODS

A total of 4,100 protocols from five different research conducted by the authors (Rosario-Hernández et al., 2013, 2015, 2018a,b, 2019, 2020) in Puerto Rico and each selected through a non-probabilistic sample, and distributed into this five large groups: sample 1 (n = 518, 12.6%), sample 2 (n = 1046, 25.5%), sample 3 (n = 1107, 27.0%), sample 4 (n = 626, 15.3%), and sample 5 (n = 803, 19.6%). The distributional differences in the five samples in sex (χ^2 [5] = 13.29, p = 0.02, Cramer's V = 0.053), level of education (Kruskall–Wallis' H [5] = 52.56, $p < 0.01$, $\chi^2 = 0.01$) and age (Kruskall–Wallis' H [5] = 74.97, $p < 0.01$, $\chi^2 = 0.01$), although they were statistically significant, the effect size was trivial, that is, $\eta^2 = 0.02$. Regarding the job characteristics of the position, the type of employment (χ^2 [5] = 28.14, $p < 0.01$, Cramer's V = 0.07), type of position (χ^2 [5] = 19.43, $p < 0.01$, Cramer's V = 0.06), type of company (χ^2 [10] = 17.02, $p < 0.01$, Cramer's V = 0.13) and years of work in the company (Kruskall–Wallis' H [5] = 369.14, $p < 0.01$, $\eta^2 = 0.08$) were not substantially different between the five samples.

The characteristic of the whole sample such as gender, age, among other, are shown in **Table 2**. The sample was composed of 56.6% of females and the average education level was 16.73 ± 2.04, which is equivalent to a bachelor's degree to one year of graduated studies.

Measures

Work-Related Rumination Scale

The WRRS was developed by Cropley et al. (2012) and has 15 questions using a 5-point Likert scale (1 = very seldom or never, 2 = seldom, 3 = sometimes, 4 = often, and 5 = very often or

TABLE 2 | Sociodemographic results of sample.

Variable	*f*	%
Gender		
Males	1,619	39.5
Females	2,320	56.6
Age (Career Stage)		
21–30 (Early Career)	1,041	25.4
31–50 (In Prime of Career)	2,235	54.5
≥51 (Past Peak of Career)	783	19.2
Position type		
Management	870	21.2
Non-management	3,083	75.2
Employment type		
Tenure	3,201	78.1
Temporary	810	19.8
Organization type		
Public state	1,253	30.6
Public federal	254	6.2
Private	2,505	61.1
Source of data		
Study 1	518	10,6
Study 2	1046	21,4
Study 3	1107	22,6
Study 4	626	12,8
Study 5	803	16,4
	Mean	***SD***
Education	16.73	2.04

Original sample size was n = 4,100.

always). According to Cropley et al. (2012) results using the factor analytic technique support a three-factor internal structure of the WRRS, which are affective rumination, problem-solving pondering, and detachment; and authors reported their reliability via Cronbach's alpha of 0.90, 0.81, and 0.88, respectively. An item example is: "Do you become tense when you think about work-related issues during your free time?

Depression

To measure depression, we used the PHQ-9 developed by Kroenke et al. (2001). The PHQ-9 is a nine-item questionnaire used for the assessment of depressive symptoms in primary care settings. This questionnaire evaluates the presence of depressive symptoms over the 2 weeks prior to the test's being filled out. Each of the items can be scored from 0 (not at all), to 3 (nearly every day). Its validity and reliability as a diagnostic measure, as well as its utility in assessing depression severity and monitoring treatment response are well established (Kroenke et al., 2001; Löwe et al., 2004a,b, 2006). In the current study, the unidimensionality of the PHQ-9 was supported by a CFA analysis using the method of robust maximum likelihood, χ^2 = 401.44(20), CFI = 0.904, SRMR = 0.047, RMSEA = 0.093[0.085;0.101]; reliability of the PHQ-9 using the omega (ω) was.899 (95% CI = 0.889;0.908.) An item example is: "Little interest or pleasure in doing things?

Anxiety

To measure anxiety, we used the GAD-7 (Spitzer et al., 2006). The GAD-7 is a seven-item questionnaire that measures general anxiety symptomatology and asked patients how often, during the last 2 weeks, they were bothered by each symptom. Response options were "not at all," "several days," "more than half the days," and "nearly every day," scored as 0, 1, 2, and 3, respectively. In addition, an item to assess duration of anxiety symptoms was included. Authors of the scale reported a Cronbach's alpha coefficient of 0.93. In terms of its construct validity, internal structure was supported by factor analysis technique and convergent validity with its association to similar measures such as the Beck Anxiety Inventory and the anxiety subscale of the Symptom Checklist-90. The unidimensionality of the GAD-7 was supported by a CFA using the robust maximum likelihood estimator, χ^2 = 154.69(14), CFI = 0.982, SRMR = 0.021, RMSEA = 0.058[0.050;0.066]; and its reliability was calculated using the omega (ω), which was 0.930 (95% CI = 0.925;0.935). An item example is: "Feeling nervous, anxious, or on edge."

Sleep Well-Being

We used the Sleep Well-Being Indicator developed by Rovira Millán and Rosario-Hernández (2018) to measure sleep well-being. This indicator is a twelve-item instrument in a Likert-frequency response format ranging from 1 (Never) to 6 (Always). This indicator has three subscales which are sleep quantity (duration), sleep quality, and consequences related to sleep. Authors report reliability through Cronbach's alpha and ranged from 0.79 to 0.86. Factor analysis results support the internal structure of three dimensions. In the current study, we used only two subscales: sleep quantity/duration and sleep quality. Thus, we examined a two-factor structure of the Sleep Well-Being Indicator using maximum likelihood robust method, χ^2 = 0.847 (1), CFI = 0.999, SRMR = 0.004, RMSEA = 0.028 [0.000;0.090]; reliability using omega (ω) = 0.776 (95% CI = 0.749;0.800) and 0.723 (95% CI = 0.687;0.754) for the sleep quantity and sleep quality subscales, respectively. An item example is: "I had trouble falling to sleep."

Burnout

We used the Maslach Burnout Inventory – General Scale (MBI-GS; Maslach et al., 1996) to measure burnout. The MBI uses a 7-point frequency scale (ranging from 0-never to 6-daily) to indicate the extent to which they experienced each item. The emotional exhaustion and cynicism have five items each and the professional efficacy six items. In this study, we used the emotional exhaustion and cynicism subscales; therefore, we tested a two-dimension model using maximum likelihood robust method, χ^2 = 454.43 (5), CFI = 0.921, SRMR = 0.004.042, RMSEA = 0.153 [0.141;0.165]; reliability was estimated using omega (ω) = 0.908 (95% CI = 0.902;0.912) and 0.791 (95% CI = 0.779;0.802) emotional exhaustion and cynicism subscales, respectively. An item example is: "I feel tired when I get up in the morning and have to face another day on the job."

Workaholism

To measure workaholism, we used the Dutch Workaholism Scale (DUWAS; Schaufeli et al., 2009) and translated into Spanish by del Líbano et al. (2010). The DUWAS is a 10-item scale which has two dimensions with 5-item each: work excessively (e.g., "I seem to be in a hurry and racing against the clock") and work compulsively (e.g., "It's important for me to work hard even when I don't enjoy what I'm doing"). Results of the CFA from the study of del Líbano et al. (2010) support the internal structure of two dimensions. In the present study, a two-factor model was supported using maximum likelihood robust method, χ^2 = 1,736 (34), CFI = 0.917, SRMR = 0.063, RMSEA = 0.114 [0.109;0.119]. Also, reliability was estimated and its 90% confidence interval using omega (ω) = 0.776 (95% CI = 0.749;0.80) and.723 (95% CI = 0.687;0.754) for the work excessively and work compulsively subscales, respectively. An item example is: "It's important for me to work hard even when I don't enjoy what I'm doing."

Social Desirability

We used the Social Desirability Scale developed by Rosario-Hernández and Rovira Millán (2002). This is a 11-item instrument in a Likert-agreement response format ranging from 1 (Totally Disagree) to 6 (Totally Agree), which pretend to measure a response bias in which people respond to a test thinking what is acceptable socially. Authors report its internal consistency through Cronbach's alpha to be 0.86, which is an excellent reliability coefficient. Factor analysis results suggest that the Social Desirability Scale internal structure has only one factor. As part of the current study, we examined the internal structure of the Social Desirability Scale using maximum likelihood robust method and results support a one factor structure as reported by its authors, χ^2 = 2,608.64 (44), CFI = 0.907, SRMR = 0.057, RMSEA = 0.115 [0.112;0.119]; also, ω reliability was was.944

(95% CI = 0.941;0.947). An item example is: "Most people have cheated on an exam, even if it was once in their lives."

Procedures

This study was approved by the Institutional Review Board of Ponce Health Sciences University (Protocol #2006040219) on June 17, 2020. Participants in all samples were selected by a convenience non-probabilistic sample method and the inclusion criteria were to be 21 years of age or older and to work at least 20 h per week. On the other hand, participants were excluded *ante hoc*, which included if they did not agree to participate voluntarily, and *post hoc* to data collection, when their scores on the WRRS were identified as outliers.

Cross-Validation Strategy

Instead of analyzing the entire sample in a single analysis action, a cross-validation strategy was applied to assess the stability of the validity parameters in the sample. This strategy considered some presuppositions. First, although the total sample would guarantee high statistical power and lower sampling error in the estimation of the parameters, the stability of the WRRS measurement model in the study samples cannot be empirically tested. Second, validation indices based on a single sample, to quantify the expected degree of cross-validation, combine the information obtained from the estimation method or fit function, together with the sample size and number of parameters (for example, AIC, BIC, ECVI; Browne and Cudeck, 1989; Whittaker and Stapleton, 2006), but direct contrast against another sample is absent where the model can be adjusted, and its replicability evaluated. Third, in the evaluation of the stability of the model where k samples drawn from the total sample are used, the cross-validation indices summarily report a discrepancy between the restricted variance-covariance matrix of the calibration sample, and the variance matrix-unconstrained covariances of the validation sample (Cudeck and Browne, 1983), but do not indicate the specific sources of the discrepancies, for example, the difference between the factor loadings in the compared samples (Byrne, 2012, p. 261). Therefore, the approach of Byrne (2012, p. 261) was followed, in which the naturally independent samples of the present study were compared within the framework of measurement invariance, and according to this, the degree of replicability of the WRRS measurement model. According to the above, the measurement model of the three oblique factors was evaluated in each subsample regarding its dimensionality, and its measurement invariance. With these two criteria met, the analysis continued toward modeling in the total sample.

Detection of Response Biases

A detection of multivariate outliers was made in the responses to all the items of the WRRS using the square Mahalanobis distance (D^2) value, an efficient and sensitive measure for outliers derived from random responses (Zijlstra et al., 2011). The cut-off point for D^2 was 3.57 (df = 15). The procedure was strengthened with the search for the longest strings of characters (long-string; Curran, 2016) based on a cut-off point (Curran, 2016): the number of consecutive repeated responses $\geq$ half the number of items ($n_{RR} \geq k/2$). The R *careless* program was used (Yentes and Wilhelm, 2018).

Item Analysis

Descriptive statistics (central response, dispersion, and distribution) and association with gender (Glass rank biserial correlation coefficient; Mangiafico, 2021) and age (ordinal eta squared; Mangiafico, 2021) were reported at the item level. The R *MVN* (Korkmaz et al., 2014) and *rcompanion* (Mangiafico, 2021) programs were used.

Internal Structure

The internal structure was evaluated through confirmatory factor analysis (CFA-SEM) and exploratory structural equation modeling (ESEM), to evaluate various measurement models of the WRRS. First, the model established by the author, consisting of three related dimensions (3F), was tested. The second model was unidimensional, to represent the use of the total score and the complete absence of discriminative validity between the dimensions, and a third model in which two-dimensional factor was tested. This third model was justified because some studies referred to a unified score for two dimensions: AR and PSP (e.g., Cropley et al., 2016, 2017; Weigelt et al., 2019a,b; Cropley and Collis, 2020). ESEM was implemented with oblique geomin target rotation (Mansolf and Reise, 2016).

In all the WRR modeling, the estimator used was WLSMV (Muthén et al., 1997) due to its effectiveness (Li, 2016), with inter-item polychoric correlations. The evaluation of the fit was made approximate fit indices (AFI): CFI ($\geq$0.95), RMSEA ($\leq$0.05), SRMR ($\leq$0.05), WRMR ($\leq$0.90; Yu, 2002). The detection of the misspecifications in the models was done with the approach of Saris et al. (2009), considering the statistical power and the size of the misspecification. Additionally, because the ESEM method estimates cross-factor loadings, the degree of factorial complexity can be observed. For this purpose, the Hoffman coefficient (Ch_{off}; Hofmann, 1977, 1978) was used; C_{hoff} values at, or near, 1.0 (Pettersson and Turkheimer, 2010), indicate that items load significantly on more than one factor (i.e., factor complexity). The modeling was carried out by the *lavaan* (Rosseel, 2012), *semtools* (Jorgensen et al., 2021), and *EFA.dimensions* (O'Connor, 2021) R programs.

Measurement invariance was done with a bottom–up approach, from an unrestricted model to a model with strong restrictions (Stark et al., 2006). Thus, we tested: an unrestricted model of equality (configurational invariance) and continued with successive restrictions applied to factor loadings and thresholds (metric invariance), and intercepts (scalar invariance). Taking into account the sample size (>300; Chen, 2007), the invariance criterion was: CFI < 0.010, SRMR < 0.030, and RMSEA < 0.015 (Chen, 2007).

Reliability Analysis

The reliability estimation was made with the coefficient ω (Green and Yang, 2009), with the method for categorical variables (Yang and Green, 2015); but since the α coefficient was usually reported in previous studies, for comparison purposes this coefficient was also estimated. Confidence intervals at 95% confidence were generated with bootstrap simulation (500 simulated samples). The precision in the direct score metric was estimated using the standard error of measurement ($SEMr_{xx}$), which should optimally be less than 0.5 (SD) to have the maximum tolerable measurement error around the observed scores (Wyrwich et al., 1999; Wyrwich, 2004). $SEMr_{xx}$ was calculated with the R program *psychometric* (Fletcher, 2010).

At the item level, reliability (r_{ii}) was estimated, which was conceptualized as the degree of response replicability in two independent applications of the item in the same participants (Zijlmans et al., 2018b, p. 999). Due to its efficacy, the classical test theory approach was used, based on the alpha coefficient as lower bound reliability, and the square of the item-test relationship (Zijlmans et al., 2018b); According to the analysis of empirical data (Zijlmans et al., 2018a), a heuristic value of $r_{ii} \geq 0.30$ is recommended as an acceptable minimum. An *ad hoc* program was used (Zijlmans et al., 2018b).

Convergent and Divergent Validity

To establish convergent and divergent validity of the WRRS, we conducted a multiple correlation analysis using observed scores via the Pearson product-moment correlation coefficient. The criterion used in this part was in two steps: first, the statistical significance set at $p < 0.01$; and second, the direction of the correlations obtained (i.e., positive, or negative). We hypothesized that AR and PSP would correlate significantly and positively with depression, anxiety, emotional exhaustion, cynicism, and workaholism; on the other hand, we expected a significantly and negative correlation with sleep duration and sleep quality. In terms of the relationship to social desirability, we expected negative and lower coefficient correlations. Meanwhile, we expected that detachment would obtain correlations significantly and negatively with depression, anxiety, emotional exhaustion, cynicism, and workaholism; on the other hand, we expected significantly and positively correlation with sleep duration and sleep quality. Regarding social desirability, we expected a negative and a lower correlation coefficient.

Descriptive Statistic and Normative Data of the Work-Related Rumination Scale

Descriptive statistics was estimated for the WRRS, such as the mean, standard deviation, standard error of measurement, possible range of scores of each factor, and the 95% confidence intervals. Normative data was produced to help interpret scores on the three factors of the WRRS.

RESULTS

Detection of Response Biases

In each of the independent samples, no more than 100 multivariate outliers were found, with a general median of 50 participants in each sample (sample 1 = 27, sample 2 = 76, sample 3 = 72, sample 4 = 52, sample 5 = 48) and $D^2 > 3.57$; altogether, 291 outliers were identified. Regarding the longest sequences of equal responses in the 15 items, according to Curran's (2016) rule, a median of 42 participants with equal responses was found in the five samples (sample 1 = 15, sample 2 = 50, sample 3 = 80, sample 4 = 81, sample 5 = 30). With both, the final effective sample for subsequent analyzes was 3576 (sample 1, n = 476, 13.3%), sample 2 (n = 921, 25.8%), sample 3 (n = 956, 26.7%), sample 4 (n = 496, 13.9%; sample 5, n = 727, 2.3%).

Item Analysis

Distribution

The multivariate normality (Henze-Zirkler's test; HZ) in the total sample was rejected (HZ test = 2.33, $p < 0.01$), as well as in the five subsamples (HZ test between 1.26 and 2.52, $p < 0.001$; see **Supplementary Tables 1–5**). There was also consistency in the absence of univariate normality in the items (SW: Shapiro–Wilk test) of the three subscales, in the clean total sample (**Table 3**), and in each of the subsamples (see **Supplementary Tables 1–5**). This was linked to the distributional skewness and excess kurtosis of the items; particularly, scale 3 showed a trend toward higher kurtosis. The similarity of the asymmetry pattern in the items was moderately high (one-way absolute agreement ICC = 0.746, 95% CI = 0.566,0.889), but the kurtosis pattern was low (one-way absolute agreement ICC = 0.227, 95% CI = 0.053,0.506); the latter suggests varied response dispersions.

Central Answer

Statistically significant differences were detected in the mean response in the total sample (**Table 3**), and in the AR factor (Friedman $\chi^2 = 1136.4$, df = 3, $p < 0.001$), PSP factor (Friedman $\chi^2 = 1049.7$, df = 4, $p < 0.001$), and Detachment factor (Friedman $\chi^2 = 27.58$, df = 2, $p < 0.001$). The size of this inter-item difference

TABLE 3 | Univariate descriptive for items in total sample (n = 3,576).

Factor/Item	M	DE	Sk	Ku	SW*	Rg_sex	Eta age
AR							
WRR1	2.62	1.26	0.3	–0.86	0.89	–0.05	0.0
WRR5	2.7	1.24	0.2	–0.93	0.9	–0.01	0.0
WRR7	2.07	1.23	0.9	–0.28	0.81	–0.02	0.0
WRR9	2.14	1.19	0.74	–0.47	0.84	0	0.0
WRR15	2.33	1.22	0.55	–0.66	0.87	–0.02	0.0
PSP							
WRR2	3.15	1.2	–0.18	–0.78	0.91	0.04	0.0
WRR4	2.97	1.15	–0.05	–0.74	0.92	0	0.0
WRR8	2.83	1.23	0.07	–0.93	0.91	0.03	0.0
WRR11	2.43	1.17	0.37	–0.72	0.89	0.08	0.0
WRR13	3.1	1.21	–0.18	–0.8	0.91	–0.02	0.0
Det							
WRR3	3.34	1.28	–0.33	–0.91	0.9	–0.01	0.0
WRR6	3.43	1.41	–0.38	–1.17	0.86	0	0.0
WRR10	3.25	1.37	–0.24	–1.15	0.89	0.01	0.0
WRR12	3.24	1.31	–0.19	–1.06	0.9	0.02	0.0
WRR14	3.48	1.24	–0.44	–0.75	0.89	0.03	0.0

*AR, affective rumination; PSP, problem-solving pondering; Det, detachment; Sk, skew coefficient; Ku, kurtosis; SW, Shapiro–Wilk test of normality; Rg, Glass' coefficient; Eta, ordinal eta coefficient. *All SW tests were significant at p < 0.01.*

can be considered large (≥0.30: large; Mangiafico, 2021) for the AR factor (Kendall W = 0.705, 95% CI:0.519,0.742), PSP factor (Kendall W = 0.579, 95% CI:0.310,0.632), and Detachment factor (Kendall W = 666, 95% CI:0.378,0.736). For each subsample, the mean response in each subscale is found in **Supplementary Tables 1–5**, where the trend seems to repeat what was found in the total sample.

Internal Structure Validity Evidence

The measurement model of the three oblique factors was evaluated in each subsample in its dimensionality, and jointly in its measurement invariance. With these two criteria fulfilled, the modeling was evaluated in the total sample. Three iterations of the modeling were made, corresponding to the evaluation of the initial dimensional structure, the process of modifying the model, and the definition of the final model, respectively.

First Iteration

In **Table 4**, the adjustment of the 15 items in each subsample is shown, with both CFA and ESEM approaches. In each sample, the fit obtained using the CFA (CFI > 0.94, RMSEA < 0.16, SRMR < 0.11, WRMR > 1.90) predominantly did not give a favorable impression to the models, since most of them deviated from the criteria to adjustment priori. In contrast, with the ESEM approach the values obtained (CFI > 0.98, RMSEA < 0.04, SRMR < 0.040, WRMR < 1.11) show a robust trend of the fit, it can be considered excellent. Additionally, the one-dimensional and two-dimensional models had a poor fit in each of the samples and in the total sample, so these models were not interpreted (see **Supplementary Table 6**).

The parameters of the factor loadings and correlations with both the ESEM and CFA approaches are shown respectively in **Supplementary Tables 7, 8**. Regarding factor loadings, these were frequently high (≥0.60) and similar within the dimensions themselves, with few exceptions. The factorial complexity of the total ESEM solution (**Supplementary Table 7**) in all the samples varied between 56 and 76% of the items, that is, more than half showed factorial complexity, that is, approximately greater than 1.5. Specifically, several items showed a consistently high degree of factorial complexity in the five subsamples; In the metric of

TABLE 4 | Fit indices of the CFA y ESEM of the WRRS Models on the five samples.

	Fit index criterion	Sample 1		Sample 2		Sample 3		Sample 4		Sample 5	
		CFA	ESEM	CFA	ESEM	CFA	ESEM	CFA	ESEM	CFA	ESEM
1st iteration											
WLSMV χ^2		653.54	49.62	1206.97	201.43	89.97	118.93	629.5	96.54	154.3	101.54
P	>0.05	<0.001	1.0	<0.001	<0.001	<0.001	0.084	<0.001	1.0	<0.001	0.41
CFI	<0.95	0.962	1.0	0.978	0.998	0.972	0.999	0.965	1.0	0.958	1.0
RMSEA	<0.05	0.117	0.00	0.118	0.034	0.098	0.15	0.112	0.00	0.152	0.00
inf		0.109	0.00	0.112	0.027	0.093	0.00	0.104	0.00	0.145	0.00
sup	<0.05	0.126	0.00	0.124	0.04	0.104	0.023	0.121	0.022	0.158	0.00
SRMR	<0.05	0.098	0.025	0.08	0.036	0.073	0.028	0.083	0.035	0.106	0.028
WRMR	<1.00	1.996	0.55	2.705	1.105	2.324	0.849	1.953	0.765	3.055	0.785
2nd iteration		**CFA**	**ESEM**	**CFA**	**ESEM**	**CFA**	**ESEM**	**CFA**	**ESEM**	**CFA**	**ESEM**
WLSMV χ^2	>0.05	–	21.403	–	7.99	–	7.884	–	54.057	–	45.77
p	<0.95	–	1.00	–	0.15	–	0.15	–	0.69	–	0.91
CFI	< 0.05	–	1.00	–	1.00	–	1.001	–	11.00	–	11.00
RMSEA		–	0.00	–	0.014	–	0.014	–	0.00	–	0.00
inf		–	0.00	–	0.00	–	0.00	–	0.00	–	0.00
sup	<0.05	–	0.00	–	0.026	–	0.026	–	0.02	–	0.009
SRMR	<0.05	–	0.022	–	0.027	–	0.027	–	0.033	–	0.023
WRMR	<1.00	–	0.433	–	0.789	–	0.789	–	0.689	–	0.634
3rd iteration		**CFA**	**ESEM**	**CFA**	**ESEM**	**CFA**	**ESEM**	**CFA**	**ESEM**	**CFA**	**ESEM**
WLSMV χ^2	>0.05	131.315	21.403	446.687	11.295	251.116	61.888	93.394	146.949	114.553	34.687
P	<0.95	<0.001	1.00	<0.001	1.00	<0.001	0.102	<0.001	<0.001	<0.001	0.939
CFI	<0.05	0.988	1.00	0.982	1.00	0.986	0.999	0.992	0.986	0.996	1.00
RMSEA		0.068	0.00	0.104	0.00	0.073	0.017	0.051	0.064	0.05	0.00
inf		0.05	0.00	0.095	0.00	0.065	0.00	0.037	0.052	0.039	0.00
sup	<0.05	0.081	0.00	0.112	0.00	0.082	0.028	0.064	0.075	0.061	0.006
SRMR	<0.05	0.063	0.022	0.069	0.017	0.056	0.028	0.049	0.06	0.043	0.022
WRMR	<1.00	1.152	0.433	2.124	0.338	1.593	0.791	0.971	1.128	1.076	

CFA, confirmatory factor analysis; ESEM, exploratory structural equation modeling.

the Hoffman coefficient, the complexity was expressed in its cross loads in two factors. These items were: 5, 6, and 13, which also showed consistently low loads or at the minimum limit ($\geq$0.50). The cross-loadings of these items were around 0.30 or more.

Regarding the inter-factor correlations, the association pattern was theoretically consistent in which a positive covariation between AR and PSP, and negative between detachment and AR and detachment and PSP. The magnitude of this covariation, however, was conditioned by the analysis approach: the estimates based on the ESEM were all attenuated (i.e., smaller in size). Taking as reference the correlations obtained with the CFA, **Supplementary Table 8** ($100(\theta_{CFA}$ - $\theta_{ESEM})/\theta_{CFA}$), the average percentage of attenuation of the interfactorial correlations with ESEM varied between 24.8 and 35.7%.

Since item reliability was one of the quality criteria of the instrument (**Supplementary Table 7**, head r_{ii}), this section also reports on this parameter. Response reproducibility through item level reliability was generally satisfactory, and most of the coefficients were > 0.40. Some items with low reliability in one sample (<0.40) showed adequate reliability in other samples and can be considered sampling error.

Second Iteration

Since the models evaluated with ESEM presented unsatisfactory specific parameters (frequent factorial complexity, and low factorial loads), and over-estimated interfactorial correlations, exclusion criteria were used based on statistical and conceptual decisions. Statistical decisions consisted of (a) the degree of factorial complexity; and (b) item level reliability should be as high as possible, at a minimum of 0.30 but with an emphasis on > 0.40. Conceptually, the exclusion criterion is the apparent redundancy of content, or the possible similar interpretation of the item chosen with another of the items of the construct. Considering these three criteria, items 5 of AR, 13 of PSP and 6 of detachment factors were eliminated. After removing these items, the ESEM was used again, but not the CFA because the decision-making was based on the ESEM results exclusively. **Supplementary Table 9** shows the adjustment of the second iteration, in which an excellent adjustment is observed, with all the indices successfully fulfilled. In the parameters obtained (factorial loads and interfactorial correlations). It is observed that the percentage of complexity of the factorial solution decreased compared to the results of the first iteration, and in each subsample the C_{hoff} median was substantially low (respectively: 1.03, 1.13, 1.08, 1.06, and 1.03); on the other hand, the factor loadings were high and moderately similar. However, item 14 was identified as potentially problematic, due to its moderate complexity in all samples, and its comparatively lower factor loading with respect to the items of its dimension. This consistency and the decision to obtain a measure with the least complexity possible, led to the removal of this item, whose content represents the behavior of the detachment factor. This item read: Do you find it easy to unwind after work?

Third Iteration

After removing item 14, the model with the remaining 11 items again fitted to the data. The ESEM fit was excellent compared to the CFA fit (**Table 4**, 3rd iteration heading), which, although it was satisfactory, was not better than the ESEM fit. In **Supplementary Table 10**, it is observed that all factorial loads were > 0.50 and predominantly were > 0.60; the complexity coefficients were close to 1.0 (except for item 11, but inconsistently in the subsamples), and the item reliability coefficients were frequently > 0.40. In contrast, the estimates produced by CFA again showed an overestimation of factor loadings and factor correlations (**Supplementary Table 11**). On the other hand, the factorial complexity (**Table 5**) was substantially lower (M = 1.04, min = 1.00, max = 1.11) compared to the complexity obtained in the previous iterations and indicated that the cross loads are predominantly considered trivial, and that the items essentially represent a single dimension. Regarding the reliability of the item of the final model, all the items exceeded the chosen criterion (>0.30), with a wide variation, but predominantly high (M = 0.47, min = 0.31, max = 0.74).

Measurement Invariance

Within Samples

The measurement invariance in every group analyzed (i.e., sex and age) was good, keeping until intercepts of scalar metric. In the **Supplementary Table 13**, the differences between fit indices (Δ_{CFI}, Δ_{RMSEA}, and Δ_{SRMR}) keeping predominantly below 0.0. In age 3 groups (**Supplementary Table 13**), the measurement invariance also was moderately satisfactory, with some changes in the consecutive models assessed, particularly in equality of intercepts model. Probably, the unbalanced sample size among the age groups in each subsample (e.g., in sample 5 one of the groups had n = 80), could have generated Type I error.

Between Samples

The number of dimensions (i.e., configurational invariance), factor loadings and thresholds (i.e., metric invariance) and latent item response (i.e., scalar invariance) were satisfactory in the five samples analyzed.

Total Sample Fit

Due to the invariance achieved between the five independent samples, the model fit of the instrument (three factors, 11 items) was estimated in the total sample (n = 3,576), in which differences conditioned by the analysis approach were again observed (**Table 5**). In the CFA approach, the adjustment was partially satisfactory because while some indicators were satisfactory (CFI = 0.989, SRMR = 0.051), other indices showed decrease (RMSEA = 0.072, 90% CI = 0.068,0.077; WRMR = 2,895); the inferential statistic was statistically significant (WLSMV – χ^2 = 809.02, p < 0.01). On the other hand, the ESEM approach produced very satisfactory results: WLSMV – χ^2 = 114.34 (p < 0.01), CFI = 0.999, RMSEA = 0.019 (90% CI = 0.015,0.024), SRMR = 0.019, and WRMR = 1.074. In the **Table 6**,

Reliability – Internal Consistency

Table 7 shows the results of the reliability estimation, with the alpha and omega coefficients. Using the standard deviation of AR (SD = 4.147, SE = 0.042), PSP (SD = 3.64, SE = 0.037) and

Detachment (SD = 3.224, SE = 0.031), the standard error of measurement (SEMr$_{xx}$) for the three WRR scores (see **Table 5**, heading SEMr$_{xx}$). According to the suggestion of Wyrwich et al. (1999), SEMr$_{xx}$ of each score was less than half the standard deviation of the score for AR and PSP, but not for Detachment (2.07, 1.82, and 1.61, respectively).

Evidence of Convergent and Divergent Validity

To gather and to establish the convergent and divergent validity of the WRRS – Spanish version, we correlated the scores of its three factors between them and to scores of others measurement instrument. **Table 8** shows that AR and PSP have a positive correlation (r = 0.478, p < 0.01) and detachment correlated negatively to AR and PSP (r = –0.329, p < 0.01, and r = –0.261, p < 0.01, respectively). Meanwhile, AR (F1) and PSP (F2) correlated positively to depression, anxiety, emotional exhaustion, cynicism, and workaholism; on the contrary, Detachment (F3) correlated negatively to those variables, as expected. On the hand, AR and PSP correlated negatively to sleep duration, sleep quality, and social desirability, whereas detachment

TABLE 5 | Factor loadings in the three factors (n_{total} = 3,576).

	CFA			ESEM				
	AR	PSP	Det	AR	PSP	Det	Rel_i	C_{hoff}
WRR1	0.809			0.722	0.078	– 0.056	0.499	1.035
WRR7	0.91			0.771	0.116	– 0.095	0.436	1.076
WRR9	0.874			0.888	– 0.012	– 0.009	0.692	1.00
WRR15	0.711			0.891	– 0.121	0.124	0.490	1.076
WRR2		0.646		– 0.05	0.687	– 0.042	0.408	1.018
WRR4		0.797		0.103	0.715	0.004	0.554	1.041
WRR8		0.819		0.123	0.678	– 0.043	0.535	1.074
WRR11		0.501		– 0.143	0.723	0.093	0.312	1.112
WRR3			0.763	– 0.096	– 0.052	0.631	0.41	1.06
WRR10			0.613	0.048	0.100	0.741	0.389	1.044
WRR12			0.864	0.027	– 0.030	0.881	0.741	1.004
% complex								29.0%
Correlation								
AR	1			1				
PSP	0.645	1		0.583	1			
Det	– 0.465	–0.400	1	– 0.427	– 0.370	1		

C_{hoff}, Hoffman's complexity index; Rel_i, reliability at the item level; CFA, confirmatory factor analysis approach; ESEM, exploratory structural equation modeling approach.

TABLE 6 | Measurement invariance in WRRS: five samples (n_{total} = 3,576).

	Fit and differences								
	χ^2	df	Δ	RMSEA	Δ	CFI	Δ	SRMR	Δ
Configurational	1037.10	205.00	–	0.10	–	0.96	–	0.06	–
Loading + thresholds	1737.50	325.00	341.49	0.09	–0.02	0.95	–0.01	0.07	0.01
Intercepts	1869.90	357.00	71.37	0.08	0.00	0.95	0.00	0.07	0.00

TABLE 7 | Internal consistency reliability.

	Reliability coefficients					
	Cronbach's Alpha (α)			*McDonalds' Omega (ω)*		
	AR	PSP	Det	AR	PSP	Det
Point estimation	0.86	0.76	0.74	0.87	0.77	0.76
SE	0.004	0.007	0.008	0.003	0.006	0.008
95% CI	(0.85,0.87)	(0.74,0.77)	(0.72,0.75)	(0.86,0.87)	(0.75,0.78)	(0.74,0.77)
SEM$_{rxx}$	1.55	1.78	1.85	1.49	1.74	1.78
Criterion: <0.5(SD)	2.07	1.82	1.61	2.07	1.82	1.61

n = 3,576; SE, standard error; SEM$_{rxx}$, standard error of measurement.

correlated positively to those variables, also as expected (see **Table 8**).

Finally, we estimated the mean, standard deviation, range, and 95% confidence interval of the WRRS-Spanish version to describe it scores. Also, we provide some guidelines to better understand and interpret the scores of WRRS (see **Table 9**).

DISCUSSION

The essential strategy of the present study was to analyze different sets of samples, obtained in different study contexts; this enhanced the inspection of the stability of the results by evaluating the measurement invariance, and in a general perspective, the replicability (de Rooij and Weeda, 2020). The estimated correlations between the latent variables based on the CFA approach were consistently different between the estimates based on the ESEM approach, to a degree it produced changes in the qualitative classification of the correlations. For example, practically all the latent correlations obtained in the CFA can be classified as high, according to the suggestions of Cohen (1992; 0.10, small; 0.30, medium; 0.50, large), or to empirically based classifications (≥0.32: 75th percentile, Bosco et al., 2015; ≥0.30: large, Gignac and Szodorai, 2016; ≥0.40, Lovakov and Agadullina, 2021). However, correlational estimates with CFA may appear to be not only high, but very high. With the ESEM, the classification of the correlational magnitude did not change, but the quantitative difference was closer to the points that separate a high magnitude from a moderate one, with the consequent impression that these correlations are high, but not very high. According to the mathematical theory behind ESEM, attenuation is produced by the estimation mechanism underlying the cross-factor loadings, in which the variance of the correlations moves toward the cross-loadings. These cross-loadings of the WRRS items are realistic representations of how the items are associated with their dimension and the rest of the dimensions, and due to the ESEM method these could be estimated. In contrast, the CFA imposes that these cross-loadings are zero, and therefore unrealistically represents the internal structure of the measurements in general, and of the WRRS in particular. Because ESEM is an approach that unites the exploratory and confirmatory approaches, the results within the exploratory framework generally carry information that leads to the analysis of factorial complexity (Fleming and Merino Soto, 2005). This result has two implications: first, that the dimensions of the WRRS maintain high correlations with each other, but not so high as to suggest a global dimension with significant interpretation; and that the correlations estimated in previous studies may be overestimated. Because the incorporation of ESEM to study the internal structure estimates the cross-loadings of the items with different factors than expected, one of the quality parameters of the internal structure was the factorial complexity, operationally defined as the degree to which the cross-loadings are different from zero. As a quality parameter, this complexity was moderately high in the first iteration of the analysis, with the full instrument as it is usually used. This highlights the consequent problem of the interpretability of the items because some of these items add invalid variance to their dimensions, because the items can represent more than one dimension. The practical implication is that, in research or professional applications, possibly a part of the contents of each dimension also incorporates other constructs of the WRRS model, to an extent that is questionable from measurement theory, that is, that a construct requires to be essentially one-dimensional to be interpreted. In the practice of construction of measures and validation, it is usual that the factorial simplicity of the items is presumed, that is, that the items purely represent their intended factors to be measured. With this conceptualization of measurement, the CFA applied to the WRRS is perfectly justified, because the cross-loadings do not exist because they are specified *a priori* with a value of zero.

In the three iterations of the ESEM analysis, the factorial complexity decreased due to the decisions made on the complex items, that is, they were removed on a statistical and substantive basis. One of the items removed was item 6 (detachment), whose responses need to be recoded to be joined to the other responses of its factor. Together with the strong magnitude of factorial complexity, its factorial loading in its expected dimension was very low; and both problems were reproducible in all five samples. It is known that the required recoding items usually produce method variance associated with their phrasing (DiStefano and Motl, 2009; Kam, 2018), and it is a problem commonly associated with the emergence of additional but spurious factors and low factor loadings. Therefore, the removal of this item, together with the rest of the removed items, produced an increase in the degree of fit of the WRRS model. A practical implication of this result for the user is that, as a first option to obtain more valid scores, remove this item from the calculation of the detachment factor

TABLE 8 | Correlation between the subscales of Work-Related Rumination Scale – Spanish version and other measures to establish convergent and divergent validity.

Scale/Variable	Affective rumination	Problem-solving pondering	Detachment
F1: Affective rumination (n = 3,576)	1		
F2: Problem-solving pondering (n = 3,576)	0.478**	1	
F3: Detachment (n = 3,576)	– 0.329**	– 0.261**	1
Depression (n = 972)	0.555**	0.209**	– 0.161**
Anxiety (n = 1,699)	0.704**	0.378**	– 0.150**
Sleep duration (n = 2,353)	– 0.314**	– 0.081**	0.223**
Sleep quality (n = 2,353)	– 0.373**	– 0.155**	0.180**
Emotional exhaustion (n = 3,100)	0.616**	0.255**	– 0.218**
Cynicism (n = 3,100)	0.540**	0.183**	– 0.138**
Workaholism (n = 727)	0.630**	0.519**	– 0.178**
Social desirability (n = 3,100)	– 0.099**	– 0.084**	0.031

*p < 0.05, **p < 0.001.

TABLE 9 | Descriptive statistics of the Work-Related Rumination Scale -Spanish version and guidelines for the interpretation of scores.

Factor	Number of items	Mean	*SD*	Possible range	95% CI	Interpretation of Scores		
						Low	Average	High
AR	4	9.19	4.15	4–20	±3	≤6	7–14	≥15
PSP	4	11.40	3.64	4–20	±3	≤7	8–14	≥15
Det	3	9.86	3.22	3–15	±3	≤6	7–12	≥13

n = 3,576. AR, affective rumination; PSP, problem-solving pondering; Det, detachment.

score; A second option is to evaluate the validity of this item, to corroborate its questionable operation, and for this the user can implement some dimensionality evaluation approach (e.g., CFA, ESEM, etc.).

Within the evaluation of the internal structure, the measurement invariance was satisfactory in the three levels evaluated (configuration, metric and thresholds, and intercepts), which helps to make comparisons according to sex, and age groups, in this study, early career (21–30 old age), in prime career (31–50 old age) and past peak career (≥51 old age), and sex. However, with respect to the age groups assessed for invariance, it is unclear whether the absence of intercept invariance (i.e., scalar) could have been produced by real differences or by the imbalance in the sample size of the samples compared in each of the five subgroups. An evaluation with different age grouping mode may be necessary to explore this with more certainty. Also, other models of equivalence assessment, including effect size, will be needed.

Our strategy to investigate measurement invariance was implemented to each of the independent subsamples ($n = 5$), and this provided an opportunity to observe the replicability of the measurement properties of the WRRS. In this last aspect, it is highlighted that the structural properties remain similar (unless, in sex and age groups), the measured parameters remain similar, given the natural variations of the administration conditions, and the variability of the individual disposition. Given that the data cleaning was antecedent to the main analyzes, in two manifestations of probable careless/insufficient effort responses (C/IE), it is possible to think about the link between the removal of the participants with IER and the measurement invariance achieved. We also observed that the difference between the five groups in the assessment of intercept invariance (i.e., scalar invariance) was larger than the cut-off points chosen and suggested by Chen (2007). This apparent lack of scalar invariance may be influenced by the chosen criteria of Chen (2007) and not be exactly appropriate for the assessment of scalar invariance. The reason is that these criteria were developed for the comparison of two groups (in our study, there were five groups), with the estimator for normally distributed continuous variables (i.e., maximum likelihood). To conclude that the invariance was not met at this level, a corroboration of the effect size of the non-invariance may be required (Nye et al., 2019).

Regarding reliability, the coefficients α and ω, the levels obtained can be considered moderately high in a general perspective and considering the interaction between the small number of items in each subscale, the sample size and the value obtained (Ponterotto and Ruckdeschel, 2007). These levels do not indicate using the WRRS for all uses, but predominantly for group applications and where decisions on individual subjects are not needed, because the coefficients are not high (i.e., 0.85 or more), the possibility of measurement error can still be considered high (Ponterotto and Ruckdeschel, 2007). The antecedent studies with the WRRS, where the interpretations are oriented toward group responses, do not conflict with this indication. On the other hand, given the similarity of the coefficients α and ω, it is assumed that some difference between the factorial loadings were trivial (Hayes and Coutts, 2020), and did not have a substantial effect on the distance between one coefficient and the other. This distance is usually associated with the degree of equality of the factorial loadings of the items, a requirement known as tau-equivalence to validate the α coefficient (Green and Yang, 2009; Hayes and Coutts, 2020). An implication of this similarity is that the estimation of internal consistency can be satisfactorily done with the coefficient α, and without requiring SEM modeling or SEM modeling approaches to estimate the coefficient ω. If the conditions of application in future uses, and the data cleaning will be effective, this implication can be induced to other contexts. Finally, given that the standard error of measurement was greater than half the standard deviation of the detachment score, it is possible that it is necessary to incorporate revision strategies of this subscale, to improve the precision of the score (Wyrwich, 2004). These strategies may require adding an item, or refining the application of the instruments, or presenting the items in an orderly manner in each content subset.

In terms of the relationship between the three factors or zero order correlations of the WRRS, they tend to be high and positive between AR and PSP and on the other hand, these two also tend to correlate negatively and somewhat medium effect size to detachment; except in one longitudinal study in which the relationships between AR and PSP were low and fluctuated between $r = 0.07$ and $r = 0.19$ (Vahle-Hinz et al., 2017). Probably one of the main concerns regarding the WRRS is whether the AR and PSP subscales can be distinguished and thus measure different constructs. Results from this brief systematic literature review is that they appear to measure two related but different constructs. For Cropley and Zijlstra (2011), emotional arousal is one of the fundamental contrasts between AR and PSP states. Psychophysiological arousal is strong in the AR state, which is detrimental to recovery, whereas the PSP state is thought to exist without psychological or physiological arousal, making it less harmful to recovery. According to Cropley and Zijlstra, AR has a negative valence, whereas problem-solving rumination has a positive valence, especially if the process of PSP results in a solution, which is supported by research that suggests that thinking about successfully completed tasks increases positive affect, self-efficacy, and well-being (Stajkovic and Luthans, 1998; Seo et al., 2004). As a result, it's likely that ruminating with a problem-solving emphasis can help with recovery at least is not that detrimental to health as AR. Moreover, Weigelt et al. (2019a) tested different models including one which has the three dimensions of the WRRS proposed by Cropley et al. (2012) and two other constructs that are also related to thinking about work such as positive and negative work reflection and results of their CFA supported that in fact, they were five different constructs.

As a final note, in the analysis it was detected that the dispersion of the responses (induced from the different kurtosis values, and low ICC), which suggests not only little redundancy among the responses, but also that the items are sensitive to individual differences in responses, and therefore the items may be interesting units of content to explore.

Regarding the limitations of the study, first, the population representativeness is not guaranteed, because the non-random selection of the samples did not corroborate the population similarity. Second, the evaluation of the measurement invariance was done by a single procedure, since different methods can produce different percentage of Type I and Type II errors, it may be required to explore the equivalence with other methods (for example, differential operation approach of items). Third, the bifactor model was not implemented, and an assessment of multidimensionality in contrast to the dimensionality of a general factor may be required (Reise, 2012; Gignac, 2016; Rodriguez et al., 2016a,b). Finally, the reliability evaluation of the stability of the scores was not implemented; to complete the evaluation of this aspect, you should study the reproducibility of the score at different points of time, using a test–retest approach.

CONCLUSION

The final version of the instrument consisted of three moderately to highly related factors, items with increased factorial simplicity, satisfactory reproducibility of the responses to the items, high reliability of internal consistency in their scores, and strong invariance between the samples.

DATA AVAILABILITY STATEMENT

Data are available upon reasonable request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Simón Carlo, Chair of the Review Institutional Review Board, Ponce Health Sciences University, Ponce, Puerto Rico. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ER-H, LR-M, and CM-S: conceptualization, methodology, writing–original draft preparation, and writing–review and editing. CM-S and ER-H: formal Analysis. ER-H and LR-M: investigation. ER-H: supervision and funding acquisition. All author: contributed to the article and approved the submitted version.

FUNDING

The project described was supported by the RCMI Program Award Number U54MD007579 from the National Institute on Minority Health and Health Disparities. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.774472/full#supplementary-material

REFERENCES

Asparouhov, T., and Muthén, B. (2009). Exploratory structural equation modeling. *Struct. Equ. Modeling* 16, 397–438. doi: 10.1080/10705510903008204

Baas, M., De Dreu, C. K. W., and Nijstad, B. A. (2008). A meta-analysis of 25 years of mood-creativity research: hedonic tone, activation, or regulatory focus? *Psychol. Bull.* 134, 779–806. doi: 10.1037/a0012815

Bisht, N. S. (2017). "Job stressors and burnout in field officers of microfinance institutions in India: role of work-related rumination," in .*Changing business environment: gamechangers, opportunities and risks*, eds N. Delener and C. Schweikert (Vienna, Austria: Global Business and Technology Association).

Blackmore, E. R., Stansfeld, S. A., Weller, I., Munce, S., Zagorski, B. M., and Stewart, D. E. (2007). Major depressive episodes and work stress: results from a national population survey. *Am. J. Public Health* 97, 2088–2093.

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., and Pierce, C. A. (2015). Correlational effect size benchmarks. *J. Appl. Psychol.* 100, 431–449.

Brotheridge, C. M., and Grandey, A. A. (2002). Emotional labor and burnout: comparing two perspectives of "people work". *J. Vocat. Behav.* 60:1739. doi: 10.1006/jvbe.2001.1815

Browne, M. W., and Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behav. Res.* 24, 445–455. doi: 10.1207/s15327906mbr2404_4

Byrne, B. M. (2012). *Structural equation modeling with Mplus: basic concepts, applications, and programming*. Milton Park: Routledge/Taylor & Francis Group.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Struct. Equ. Modeling* 14, 464–504. doi: 10.1080/10705510701301834

Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155

Cropley, M., and Collis, H. (2020). The Association Between Work-Related Rumination and Executive Function Using the Behavior Rating Inventory of Executive Function. *Front. Psychol.* 11:821. doi: 10.3389/fpsyg.2020.00821

Cropley, M., Dijk, D.-J., and Stanley, N. (2006). Job strain, work rumination, and sleep in school teachers. *Eur. J. Work Organ. Psychol.* 15, 181–196. doi: 10.1080/13594320500513913

Cropley, M., Michalianou, G., Pravettoni, G., and Millward, L. J. (2012). The relation of post-work ruminative thinking with eating behaviour. *Stress Health* 28, 23–30. doi: 10.1002/smi.1397

Cropley, M., Plans, D., Morelli, D., Sütterlin, S., Inceoglu, I., Thomas, G., et al. (2017). The Association between Work-Related Rumination and Heart Rate Variability: a Field Study. *Front. Hum. Neurosci.* 11:27. doi: 10.3389/fnhum.2017.00027

Cropley, M., Zijlstra, F. R., Querstret, D., and Beck, S. (2016). Is Work-Related Rumination Associated with Deficits in Executive Functioning? *Front. Psychol.* 7:1524. doi: 10.3389/fpsyg.2016.01524

Cropley, M., and Zijlstra, F. R. H. (2011). "Work and rumination," in *Handbook of stress in the occupations*, eds J. Langan-Fox and C. L. Cooper (Cheltenham: Edward Elgar Publishing), 487–501. doi: 10.4337/9780857931153.00061

Crutzen, R., and Peters, G.-J. Y. (2017). Scale quality: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychol. Rev.* 11, 242–247. doi: 10.1080/17437199.2015.1124240

Cudeck, R., and Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behav. Res.* 18, 147–167. doi: 10.1207/s15327906mbr1802_2

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *J. Exp. Soc. Psychol.* 66, 4–19.

de Rooij, M., and Weeda, W. (2020). Cross-Validation: a method every psychologist should know. *Adv. Methods Pract. Psychol. Sci.* 3, 248–263.

del Líbano, M., Llorens, S., Salanova, M., and Schaufeli, W. (2010). Validity of a brief workaholism scale. *Psicothema* 22, 143–150.

DiStefano, C., and Motl, R. W. (2009). Self-esteem and method effects associated with negatively worded items: investigating factorial invariance by sex. *Struct. Equ. Modeling* 16, 134–146. doi: 10.1080/10705510802565403

Dormann, C., and Zapf, D. (1999). Social support, social stressors at work, and depressive symptoms: testing for main and moderating effects with structural equations in a three-wave longitudinal study. *J. Appl. Psychol.* 84, 874–884. doi: 10.1037/0021-9010.84.6.874

Dunn, J. M., and Sensky, T. (2018). Psychological processes in chronic embitterment: the potential contribution of rumination. *Psychol. Trauma* 10, 7–13. doi: 10.1037/tra0000291

Etzion, D., Eden, D., and Lapidot, Y. (1998). Relief from job stressors and burnout: reserve service as a respite. *J. Appl. Psychol.* 83, 577–585. doi: 10.1037/0021-9010.83.4.577

Firoozabadi, A., Uitdewilligen, S., and Zijlstra, F. R. H. (2018a). Should you switch off or stay engaged? The consequences of thinking about work on the trajectory of psychological well-being over time. *J. Occup. Health Psychol.* 23, 278–288. doi: 10.1037/ocp0000068

Firoozabadi, A., Uitdewilligen, S., and Zijlstra, F. R. H. (2018b). Solving problems or seeing troubles? A day-level study on the consequences of thinking about work on recovery and well-being, and the moderating role of self-regulation. *Eur. J. Work Organ. Psychol.* 27, 629–641. doi: 10.1080/1359432X.2018.1505720

Fleming, J. S., and Merino Soto, C. (2005). Medidas de simplicidad y de ajuste factorial: un enfoque para la evaluación de escalas construidas factorialmente. *Rev. Psicol.* 23, 250–266.

Fletcher, T. D. (2010). *Psychometric: applied Psychometric Theory. R package version 2.2.* Available Online at: https://CRAN.R-project.org/package=psychometric. (accessed February 15, 2021).

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Adv. Methods Pract. Psychol. Sci.* 3, 484–501. doi: 10.1177/2515245920951747

Fritz, C., Sonnentag, S., Spector, P. E., and McInroe, J. A. (2010). The weekend matters: relationships between stress recovery and affective experiences. *J. Organ. Behav.* 31, 1137–1162. doi: 10.1002/job.672

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: that is why the bifactor model tends to fit better. *Intelligence* 55, 57–68. doi: 10.1016/j.intell.2016.01.006

Gignac, G. E., and Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Pers. Individ. Dif.* 102, 74–78.

Grant, M. J., and Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info. Libr. J.* 26, 91–108. doi: 10.1111/j.1471-1842.2009.00848.x

Green, S. B., and Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: an alternative to coefficient alpha. *Psychometrika* 74, 155–167. doi: 10.1007/s11336-008-9099-3

Hayes, A. F., and Coutts, J. J. (2020). Use omega rather than cronbach's alpha for estimating reliability. But. *Commun. Methods Meas.* 14, 1–24. doi: 10.1080/19312458.2020.1718629

Hofmann, R. J. (1977). Indices Descriptive of Factor Complexity. *J. Gen. Psychol.* 96, 103–110. doi: 10.1080/00221309.1977.9920803

Hofmann, R. J. (1978). Complexity and simplicity as objective indices descriptive of factor solutions. *Multivariate Behav. Res.* 13, 247–225. doi: 10.1207/s15327906mbr1302_9

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., and Rosseel, Y. (2021). *semTools: useful tools for structural equation modeling. R package version .5-3. Recuperado de.* Available Online at: https://CRAN.R-project.org/package=semTools. (accessed February 15, 2021).

Junker, N. M., Baumeister, R. F., Straub, K., and Greenhaus, J. H. (2020). When forgetting what happened at work matters: the role of affective rumination, problem-solving pondering, and self-control in work-family conflict and enrichment. *J. Appl. Psychol.* doi: 10.1037/apl0000847 [Epub Online ahead of print].

Kam, C. C. S. (2018). Why do we still have an impoverished understanding of the item wording effect? An empirical examination. *Sociol. Methods Res.* 47, 574–597. doi: 10.1177/0049124115626177

Karasek, R., Baker, D., Marxer, F., Ahlbom, A., and Theorell, T. (1981). Job decision latitude, job demands, and cardiovascular disease: a prospective study of Swedish men. *Am. J. Public Health* 71, 694–705. doi: 10.2105/ajph.71.7.694

Kinnunen, U., Feldt, T., and de Bloom, J. (2019). Testing cross-lagged relationships between work-related rumination and well-being at work in a three-wave longitudinal study across 1 and 2 years. *J. Occup. Organ. Psychol.* 92, 645–670. doi: 10.1111/joop.12256

Kinnunen, U., Feldt, T., de Bloom, J., Sianoja, M., Korpela, K., and Geurts, S. (2017). Linking boundary crossing from work to nonwork to work-related rumination across time: a variable- and person-oriented approach. *J. Occup. Health Psychol.* 22, 467–480. doi: 10.1037/ocp0000037

Kivimäki, M., Virtanen, M., Elovainio, M., Kouvonen, A., Väänänen, A., and Vahtera, J. (2006). Work stress in the etiology of coronary heart disease: a meta-analysis. *Scand. J. Work Environ. Health* 32, 431–442. doi: 10.5271/sjweh.1049

Korkmaz, S., Goksuluk, D., and Zararsiz, G. (2014). MVN: an R Package for Assessing Multivariate Normality. *R J.* 6, 151–162.

Kroenke, K., Spitzer, R. L., and Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16, 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* 48, 936–949.

Lovakov, A., and Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *Eur. J. Soc. Psychol.* 51, 485–504.

Löwe, B., Kroenke, K., Herzog, W., and Gräfe, K. (2004a). Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *J. Affect. Disord.* 81, 61–66. doi: 10.1016/S0165-0327(03)00198-8

Löwe, B., Unützer, J., Callahan, C. M., Perkins, A. J., and Kroenke, K. (2004b). Monitoring depression treatment outcomes with the patient health questionnaire-9. *Med. Care* 42, 1194–1201.

Löwe, B., Schenkel, I., Carney-Doebbeling, C., and Göbel, C. (2006). Responsiveness of the PHQ-9 to Psychopharmacological Depression Treatment. *Psychosomatics* 47, 62–67. doi: 10.1176/appi.psy.47.1.62

Magnavita, N., and Fileni, A. (2014). Work stress and metabolic syndrome in radiologists: first evidence. *Radiol. Med.* 119, 142–148. doi: 10.1007/s11547-013-0329-0

Mangiafico, S. (2021). *rcompanion: functions to Support Extension Education Program Evaluation. R package version 2.4.1*. Available Online at: https://CRAN.R-project.org/package=rcompanion. (accessed February 15, 2021).

Mansolf, M., and Reise, S. P. (2016). Exploratory bifactor analysis: the Schmid-Leiman orthogonalization and Jennrich-Bentler analytic rotations. *Multivariate Behav. Res.* 51, 698–717.

Maslach, C., Jackson, S. E., and Leiter, M. P. (1996). *Maslach Burnout Inventory manual*, 3rd Edn. Palo Alto: Consulting Psychologists Press.

Mehmood, Q., and Hamstra, M. R. W. (2021). Panacea or mixed blessing? Learning goal orientation reduces psychological detachment via problem-solving rumination. *Appl. Psychol.* 70, 1841–1855. doi: 10.1111/apps.12294

Meijman, T. F., and Mulder, G. (1998). "Psychological aspects of workload," in *Handbook of work and organizational: work psychology*, eds P. J. D. Drenth, H. Thierry, and C. J. de Wolff (United Kingdom: Psychology Press/Erlbaum, Taylor & Francis), 5–33.

Merino-Soto, C., and Angulo-Ramos, M. (2020). Validity induction: comments on the study of Compliance Questionnaire for Rheumatology. *Rev. Colombiana de Reumatol.* 28, 312–313. doi: 10.1016/j.rcreu.2020.05.005

Merino-Soto, C., and Angulo-Ramos, M. (2021). Metric studies of the Compliance Questionnaire for Rheumatology (CQR): a case of validity induction? *Reumatol. Clín.* doi: 10.1016/j.reuma.2021.03.004 [Epub Online ahead of print].

Merino-Soto, C., and Calderón-De la Cruz, G. A. (2018). Validez de estudios peruanos sobre estrés y burnout [Validity of Peruvian studies on stress and burnout]. *Rev. Per. Med. Exp. Salud Pública* 35, 353–354. doi: 10.17843/rpmesp.2018.353.3521

Morin, A. J. S., Arens, A. K., and Marsh, H. W. (2015). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Struct. Equ. Modeling* 23, 116–139.

Mullen, P. R., Backer, A., Chae, N., and Li, H. (2020). School counselors' work-related rumination as predictor of burnout, turnover intention, job satisfaction, and work engagement. *Prof Sch. Couns.* 24, 1–10.

Muthén, B. O., du Toit, S., and Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Unpublished manuscript*. Available Online at: https://www.statmodel.com/download/Article_075.pdf (accessed February 15, 2021).

Nye, C. D., Bradburn, J., Olenick, J., Bialko, C., and Drasgow, F. (2019). How big are my effects? Examining the magnitude of effect sizes in studies of measurement equivalence. *Organ. Res. Methods* 22, 678–709. doi: 10.1177/1094428118761122

O'Connor, B. P. (2021). *EFA.dimensions: exploratory Factor Analysis Functions for Assessing Dimensionality. R package version .1.7.2*. Available Online at: https://CRAN.R-project.org/package=EFA.dimensions. (accessed February 15, 2021).

Pauli, R., and Lang, J. (2021). Collective resources for individual recovery: the moderating role of social climate on the relationship between job stressors and work-related rumination: a multilevel approach. *Ger. J. Hum. Resour. Manage.* 35, 152–175.

Pereira, D., and Elfering, A. (2014). Social stressors at work and sleep during weekends: the mediating role of psychological detachment. *J. Occup. Health Psychol.* 19, 85–95. doi: 10.1037/a0034928

Pettersson, E., and Turkheimer, E. (2010). Item selection, evaluation, and simple structure in personality data. *J. Res. Pers.* 44, 407–442.

Pisanti, R., Gagliardi, M. P., Razzino, S., and Bertini, M. (2003). Occupational stress and wellness among Italian secondary school teachers. *Psychol. Health* 18, 523–536. doi: 10.1080/0887044031000147247

Ponterotto, J. G., and Ruckdeschel, D. E. (2007). An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Percept. Mot. Skills* 105, 997–1014. doi: 10.2466/pms.105.3.997-1014

Pravettoni, G., Cropley, M., Leotta, S. N., and Bagnara, S. (2007). The differential role of mental rumination among industrial and knowledge workers. *Ergonomics* 50, 1931–1940. doi: 10.1080/00140130701676088

Querstret, D., and Cropley, M. (2012). Exploring the relationship between work-related rumination, sleep quality, and work-related fatigue. *J. Occup. Health Psychol.* 17, 341–353. doi: 10.1037/a0028552

Querstret, D., Cropley, M., and Fife-Schaw, C. (2017). Internet-based instructor-led mindfulness for work-related rumination, fatique, and sleep: assessing facets of mindfulness as mecahnisms of change. A randomized waitlist control trial. *J. Occup. Health Psychol.* 22, 153–169. doi: 10.1037/ocp0000028

Querstret, D., Cropley, M., Kruger, P., and Heron, R. (2016). Assessing the effect of a Cognitive Behaviour Therapy (CBT)-based workshop on work-related rumination, fatigue, and sleep. *Eur. J. Work Organ. Psychol.* 25, 50–67. doi: 10.1080/1359432X.2015.1015516

Reise, S. P. (2012). Invited paper: the rediscovery of bifactor measurement models. *Multivariate Behav. Res.* 47, 667–696.

Rodriguez, A., Reise, S. P., and Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *J. Pers. Assess.* 98, 223–237.

Rodriguez, A., Reise, S. P., and Haviland, M. G. (2016b). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychol. Methods* 21, 137–115.

Roger, D., and Jamieson, J. (1988). Individual differences in delayed heart-rate recovery following stress: the role of extraversion, neuroticism and emotional control. *Pers. Individ. Dif.* 9, 721–726. doi: 10.1016/0191-8869(88)90061-X

Rook, J. W., and Zijlstra, F. R. H. (2006). The contribution of various types of activities to recovery. *Eur. J. Work Organ. Psychol.* 15, 218–240. doi: 10.1080/13594320500513962

Rosario-Hernández, E., and Rovira Millán, L. V. (2002). Desarrollo y validación de una escala para medir las actitudes hacia el retiro. *Rev. Puertorriqueña Psicol.* 13, 45–60.

Rosario-Hernández, E., Rovira Millán, L. V., Comas Nazario, A. R., Medina Hernández, A., Colón Jiménez, R., Feliciano Rivera, Y., et al. (2018a). Workplace bullying and its effect on sleep well-being: the mediatin role of rumination. *Rev. Puertorriqueña Psicol.* 29, 164–186.

Rosario-Hernández, E., Rovira Millán, L. V., Vélez Ramos, J., Cruz, M., Vélez, E., Torres, G., et al. (2018b). Effect of the exposure to workplace bullying on turnover intention and the mediating role of job satisfaction, work engagement, and burnout. *Rev. Interamericana Psicol. Ocupacional* 37, 26–51. doi: 10.21772/ripo.v37n1a03

Rosario-Hernández, E., Rovira Millán, L. V., Díaz Pla, L., Segarra Colondres, C., Soto Franceschii, J. A., Rodríguez Irizarry, A., et al. (2013). Las demandas laborales y su relación con el bienestar psicológico y físico: el papel mediador de la rumiación relacionada con el trabajo. *Rev. Interamericana Psicol. Ocupacional* 32, 69–95.

Rosario-Hernández, E., Rovira Millán, L. V., Díaz Pla, L., Segarra Colondres, C., Soto Franceschini, J. A., Rodríguez Irizarry, A., et al. (2015). Las demandas laborales y sus efectos en el bienestar del suenno: el papel mediador de la rumiación relacionada con el trabajo. *Rev. Puertorriqueña Psicol.* 26, 150–169.

Rosario-Hernández, E., Rovira Millán, L. V., Rodríguez Irizarry, A., Rivera Alicea, B. E., Fernández López, L. N., López Miranda, R. S., et al. (2014). Salud cardiovascular y su relación con los factores de riesgo psicosociales en una muestra de empleados puertorriqueños. *Rev. Puertorriqueña Psicol.* 25, 98–116.

Rosario-Hernández, E., Rovira Millán, L. V., Sánchez-García, N. C., Padovani Rivera, C. M., Velázquez Lugo, A., Maldonado Fonseca, I. M., et al. (2020). A boring story about work: do bored employees ruminate? *Rev. Puertorriqueña Psicol.* 31, 92–108.

Rosario-Hernández, E., Rovira Millán, L. V., Vega Vélez, S., Zeno-Santi, R., Farinacci García, P., Centeno Quintana, L., et al. (2019). Exposure to workplace bullying and suicidal ideation: an exploratory study. *J. Appl. Struct. Equ. Modeling* 3, 55–75.

Rosseel, Y. (2012). lavaan: an R Package for Structural Equation Modeling. *J. Stat. Softw.* 48, 1–36.

Rovira Millán, L. V., and Rosario-Hernández, E. (2018). Desarrollo y validación del Indicador de Bienestar del Sueño. *Rev. Puertorriqueña Psicol.* 29, 348–362.

Safstrom, M., and Harting, T. (2013). Psychological detachment in the relationship between job stressors and strain. *Behav. Sci.* 3, 418–433. doi: 10.3390/bs3030418

Saris, W. E., Satorra, A., and van der Veld, W. M. (2009). Testing Structural Equation Models or Detection of Misspecifications? Structural Equation Modeling. *Struct. Equ. Modeling* 16, 561–582.

Schaufeli, W. B., Shimazu, A., and Taris, T. W. (2009). Being driven to work excessively hard: the evaluation of a two-factor measure of workaholism in the Netherlands and Japan. *Cross Cult. Res.* 43, 320–348. doi: 10.1177/1069397109337239

Schwartz, A. R., Gerin, W., Davidson, K. W., Pickering, T. G., Brosschot, J. F., Thayer, J. F., et al. (2003). Toward a causal model of cardiovascular responses to stress and the development of cardiovascular disease. *Psychosom. Med.* 65, 22–35. doi: 10.1097/01.psy.0000046075.79922.61

Seo, M., Barrett, L., and Bartunek, J. M. (2004). The role of affective experience in work motivation. *Acad. Manag. Rev.* 29, 423–439. doi: 10.5465/amr.2004.13670972

Smyth, A., de Bloom, J., Syrek, C., Domin, M., Janneck, M., Reins, J. A., et al. (2020). Efficacy of a smartphone-based intervention – "Holidaily" – promoting recovery behaviour in workers after a vacation: study protocol for randomized controlled trial. *BMC Pub. Health* 20:1286. doi: 10.1186/s12889-020-09354-5

Sonnentag, S., Mojza, E. J., Binnewies, C., and Scholl, A. (2008). Being engaged at work and detached at home: a week-level study on work engagement, psychological detachment, and affect. *Work Stress* 22, 257–276. doi: 10.1080/02678370802379440

Sonnentag, S., and Zijlstra, F. R. H. (2006). Job characteristics and off-job activities as predictors of need for recovery, well-being, and fatigue. *J. Appl. Psychol.* 91, 330–350. doi: 10.1037/0021-9010.91.2.330

Spitzer, R. L., Kroenke, K., Williams, J. B. W., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder the GAD-7. *Arch. Intern. Med.* 166, 1092–1097.

Stajkovic, A. D., and Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychol. Bull.* 124, 240–261. doi: 10.1037/0033-2909.124.2.240

Stark, S., Chernyshenko, O. S., and Drasgow, F. (2006). Detecting dierential item functioning with conrmatory factor analysis and item response theory: toward a unifed strategy. *J. Appl. Psychol.* 91, 1292–1306.

Sulak Akyüz, B., and Sulak, S. (2019). Adaptation of Work-Related Rumination Scale into Turkish. *J. Meas. Eval. Educ. Psychol.* 10, 422–434.

Svetieva, E., Clerkin, C., and Ruderman, M. N. (2017). Can't sleep, won't sleep: exploring leaders' sleep patterns, problems, and attitudes. *Consult. Psychol. J. Pract. Res.* 69, 80–97. doi: 10.1037/cpb0000092

Syrek, C. J., Weigelt, O., Peifer, C., and Antoni, C. H. (2017). Zeigarnik's sleepless nights: how unfinished tasks at the end of the week impair employee sleep on the weekend through rumination. *J. Occup. Health Psychol.* 22, 225–238. doi: 10.1037/ocp0000031

Vahle-Hinz, T., Mauno, S., de Bloom, J., and Kinnunen, U. (2017). Rumination for innovation? Analysing the longitudinal effects of work-related rumination on creativity at work and off-job recovery. *Work Stress* 31, 315–337. doi: 10.1080/02678373.2017.1303761

van der Doef, M., Mbazzi, F. B., and Verhoeven, C. (2012). Job conditions, job satisfaction, somatic complaints and burnout among East African nurses. *J. Clin. Nurs.* 21, 1763–1775. doi: 10.1111/j.1365-2702.2011.03995.x

Van Laethem, M., Beckers, D. G. J., de Bloom, J., Sianoja, M., and Kinnunen, U. (2019). Challenge and hindrance demands in relation to self-reported job performance and the role of restoration, sleep quality, and affective rumination. *J. Occup. Organ. Psychol.* 92, 225–254. doi: 10.1111/joop.12239

Vandevala, T., Pavey, L., Chelidoni, O., Chang, N. F., Creagh-Brown, B., and Cox, A. (2017). Psychological rumination and recovery from work in intensive care professionals: associations with stress, burnout, depression, and health. *J. Intensive Care* 5:16. doi: 10.1186/s40560-017-0209-0

Vassar, M., Ridge, J., and Hill, A. (2008). Inducing Score Reliability from Previous Reports: an Examination of Life Satisfaction Studies. *Soc. Indic. Res.* 87, 27–45. doi: 10.1007/s11205-007-9157-8

Weigelt, O., Gierer, P., and Syrek, C. J. (2019a). My Mind is Working Overtime—Towards an Integrative Perspective of Psychological Detachment, Work-Related Rumination, and Work Reflection. *Int. J. Environ. Res. Public Health* 16:2987. doi: 10.3390/ijerph16162987

Weigelt, O., Syrek, C. J., Schmitt, A., and Urbach, T. (2019b). Finding peace of mind when there still is so much left undone: a diary study on how job stress, competence need satisfaction, and proactive work behavior contribute to work-related rumination during the week. *J. Occup. Health Psychol.* 24, 373–386.

Whittaker, T. A., and Stapleton, L. M. (2006). The Performance of Cross-Validation Indices Used to Select Among Competing Covariance Structure Models Under Multivariate Nonnormality Conditions. *Multivariate Behav. Res.* 41, 295–335.

Wyrwich, K. W. (2004). Minimal important difference thresholds and the standard error of measurement: is there a connection? *J. Biopharm. Stat.* 14, 97–110. doi: 10.1081/bip-120028508

Wyrwich, K. W., Tierney, W. M., and Wolinsky, F. D. (1999). Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J. Clin. Epidemiol.* 52, 861–873. doi: 10.1016/S0895-4356(99)00071-2

Yang, Y., and Green, S. B. (2015). Evaluation of structural equation modeling estimates of reliability for scales with ordered categorical items. *Methodology* 11, 23–34. doi: 10.1027/1614-2241/a000087

Yentes, R. D., and Wilhelm, F. (2018). *Careless: procedures for computing indices of careless responding. R package version 1.1.3. Recuperado de.* Available Online at: https://cran.r-project.org/web/packages/careless/index.html. (accessed February 15, 2021).

Yu, C. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Ph.D. thesis. Los Angeles, CA: University of California.

Zhang, J., Li, W., Ma, H., and Smith, A. P. (2020). Switch Off Totally or Switch Off Strategically? The Consequences of Thinking About Work on Job Performance. *Psychol. Rep.* doi: 10.1177/0033294120968080 [Epub Online ahead of print].

Zijlmans, E. A. O., Tijmstra, J., van der Ark, L. A., and Sijtsma, K. (2019). Item-Score Reliability as a Selection Tool in Test Construction. *Front. Psychol.* 9:2298. doi: 10.3389/fpsyg.2018.02298

Zijlmans, E. A. O., Van der Ark, L. A., Tijmstra, J., and Sijtsma, K. (2018b). Methods for estimating item-score reliability. *Appl. Psychol. Meas.* 42, 553–557.

Zijlmans, E. A. O., Tijmstra, J., van der Ark, L. A., and Sijtsma, K. (2018a). Item-score reliability in empirical-data sets and its relationship with other item indices. *Educ. Psychol. Meas.* 78, 998–102.

Zijlstra, F. R. H., and Sonnentag, S. (2006). After work is done: psychological perspectives on recovery from work. *Eur. J. Work Organ. Psychol.* 15, 129–138. doi: 10.1080/13594320500513855

Zijlstra, W. P., van der Ark, L. A., and Sijtsma, K. (2011). Outliers in test and questionnaire data: can they be detected and should they be removed? *J. Educ. Behav. Stat.* 36, 186–212.

Zoupanou, Z., Cropley, M., and Rydstedt, L. W. (2013). Recovery after work: the role of work beliefs in the unwinding process. *PLoS One* 8:e81381. doi: 10.1371/journal.pone.0081381

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

BRIEF RESEARCH REPORT
published: 01 December 2021
doi: 10.3389/fpsyg.2021.727499

Development and Evidence of the Validity of the Condom Use Attitudes Scale for Youth and Adults in a Chilean Context

Rodrigo Ferrer-Urbina, Patricio Mena-Chamorro, Geraldy Sepúlveda-Páez and Marcos Carmona-Halty*

Escuela de Psicología y Filosofía, Universidad de Tarapacá, Arica, Chile

Condom use is the most effective preventive behavior against HIV transmission, and its inadequate use is a public health problem that occurs mostly among youth and young adults. Although there are scales that measure condom use, those that exist correspond to English-speaking developments or do not have psychometric evidence to support them, so it is possible that the available adaptations of instruments do not adequately reflect the phenomenon in the Chilean population. Thus, the study aims to develop a scale to assess attitudes toward condom use in Chilean youth and young adults. Initially, a sample of students between 18 and 39 years (n = 520) was used for debugging the instrument. Then, a second sample was taken from the general population aged 18 to 40 (n = 992) to confirm the factor structure of the proposed model. The final scale has 10 items and 3 attitudinal dimensions (affective, cognitive, and behavioral). The results show that the identified structure provides adequate levels ($\omega > 0.7$) or at least sufficient of reliability ($\omega > 0.6$) and presents evidence of validity, based on the internal structure of the test, through ESEM (CFI = 0.993; TLI = 0.984; RMSEA = 0.056). In addition, evidence of validity was obtained based on the relationship with other variables and strong invariance between the scores of men and women. It is concluded that the scale developed has adequate psychometric properties to assess, in brief form, condom use attitudes in equal samples for research and screening purposes.

Keywords: attitudes toward condom use, HIV/AIDS, sexual risk behavior, ESEM, Psychometric scales development

OPEN ACCESS

Edited by:
Antonio De Padua Serafim,
Universidade de São Paulo, Brazil

Reviewed by:
Irene Checa Esquiva,
University of Valencia, Spain
Luis Eduardo Garrido,
Pontificia Universidad Católica Madre y Maestra, Dominican Republic

****Correspondence:***
Rodrigo Ferrer-Urbina
rferrer@academicos.uta.cl

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *18 June 2021*
Accepted: *03 November 2021*
Published: *01 December 2021*

Citation:
Ferrer-Urbina R, Mena-Chamorro P, Sepúlveda-Páez G and Carmona-Halty M (2021) Development and Evidence of the Validity of the Condom Use Attitudes Scale for Youth and Adults in a Chilean Context. Front. Psychol. 12:727499. doi: 10.3389/fpsyg.2021.727499

INTRODUCTION

Human immunodeficiency virus and acquired immunodeficiency syndrome (hereafter HIV/AIDS) is a problem that affects millions of people, mainly in low- and middle-income countries (World Health Organization, 2016). Notwithstanding global efforts to prevent its transmission, the number of carriers continues to increase, with 120,000 new cases registered during 2019 in Latin America alone (Joint United Nations Program on HIV/AIDS, 2019; Pan American Health Organization, 2019).

In Chile, the context is similar, showing a systematic increase in the number of HIV/AIDS reports, with the highest prevalence figures in the regions of Arica and Parinacota, Metropolitan and Tarapacá (Cáceres-Burton, 2019). The main affected people are between 20 and 29 years, being 40.4% of the total number of new cases confirmed during 2017 (Instituto de Salud Pública, 2016; Barrera-Herrera and Vinet, 2017; Cáceres and Pino, 2018; Cáceres-Burton, 2019).

This scenario has prompted multiple governmental efforts to reduce sexual risk behaviors (hereafter, SRB) (i.e., sex with inadequate condom use, sex under the influence of alcohol and drug use, and multiple sexual partners), because they are the main way of transmission (Kilwein and Looby, 2018; Yi et al., 2018). Despite these efforts, the results of prevention interventions and programs remain insufficient, especially in the heterosexual population (Kilwein et al., 2017; Habel et al., 2018).

Among the HIV/AIDS prevention's mechanisms, the one with most support and consensus is the use of condoms (Johnson et al., 2018), since they provide an impermeable barrier for sperm-sized particles and HIV and STI pathogens, making their systematic and correct use effective for prevention (Smith et al., 2015). However, the literature points out that a high proportion of young people report not using condoms consistently, which increases the risk of HIV/STI infection (Pinyaphong et al., 2018). For example, in the Chilean context, only 22.1% of young people between 20 and 24 years and 11.5% of adults between 25 and 29 years report always using female or male condoms during the last 12 months (Ministerio de Salud, 2017).

Evidence shown that there are multiple psychosocial factors that directly or indirectly influence potential condom use, including which sexual communication (Javier et al., 2018), perceived risk (Elshiekh et al., 2020), type of sexual partner (formal or informal) (Bryan et al., 2017), previous condom use, knowledge, self-efficacy, and attitude toward condom use (Sheeran et al., 2016; Janulis et al., 2017; Teye-Kwadjo et al., 2017). While all these characteristics impact condom use, the factor that has evidenced the most significant impact on condom use is attitude (Sheeran et al., 1999; Mbelle et al., 2018).

Favorable attitudes toward condoms have been shown to lead to healthy sexual behaviors (De Torres, 2020; Kim et al., 2021), while unfavorable attitudes decrease the likelihood of condom use (Ajayi et al., 2019; Elshiekh et al., 2020) and are associated with other sexual risk behaviors, such as multiple sexual partners (Shamu et al., 2020) and sex under the influence of alcohol or drugs (Davis et al., 2020).

Although there are several instruments to measure attitudes toward condom use (e.g., Brown, 1984; Sacco et al., 1991; Helweg-Larsen and Collins, 1994; Hanna, 1999; Neilands and Choi, 2002; Crosby et al., 2010; Hollub et al., 2011), and some of these have been adapted for Spanish-speaking populations (Vallejo-Medina et al., 2019; Plaza-Vidal et al., 2020), they have some characteristics (e.g., scales specific to a social group; scales specific to a female or male condom; scales developed in a foreign language) (Beachy et al., 2020) which reduce their general usability and threaten the validity of their interpretations in particular settings.

Therefore, evidence shown that cultural differences can affect people's perceptions and attitudes toward condoms (Mileti et al., 2018; Patterson, 2019), and there are no instruments with evidence of validity to support their uses and interpretations in chilean population, according to current standards (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014). This study aims to develop a scale to measure the attitude toward condom use, with evidence of validity for its use in Chilean youth and young adults.

Considering that multiple attitudinal assessment models have been proposed in the literature, such as the theory of reasoned action (TRA; Fishbein and Ajzen, 1975) and planned action (TPB; Ajzen and Madden, 1986) models, given that we intend to develop a brief scale focused on past behavior, we opted to use a simpler attitudinal model, specifically the 3-component model: (1) Cognitive component: thoughts and beliefs held about the object; (2) Affective, referring to the affective/emotional evaluation held about the object; (3) and Behavioral, referring to the way of acting that allows making evaluations about the attitudinal object (Zanna and Rempel, 1988). Furthermore, given that some authors have argued that attitude can be analyze as a global construct (e.g., La Trobe and Acott, 2000; Gaborieau and Pronello, 2021), it was decided to contrast this possibility as an alternative one-dimensional model, to test the support to a general factor.

MATERIALS AND METHODS

An instrumental study with a cross-sectional design was conducted (Ato et al., 2013).

Participants

Two samples were used for this study: (a) one of university students between 18 and 39 years, and (b) one of the general population between 18 and 40 years. The participants of both samples were chosen using non-probability sampling strategies (Otzen and Manterola, 2017). Sample (a), collected using a time-space strategy, is composed of 520 young adults (M = 22.7; SD = 3.5) from the city of Arica, 264 (50.7%) were female and 253 (48.6%) were male, where 84.2% (n = 438) reported being heterosexual and 61.3% (n = 320) reported having used protective barrier methods in the last 2 years; sample (b), collected using a quota strategy (i.e. sex, educational level and city), according to the baseline proportions provided by the 2017 CENSUS results (Instituto Nacional de Estadística, 2018), is composed of 992 young adults (M = 23. 3; SD = 4.6), 514 (52.0%) females and 464 (46.7%) males, from the cities of Arica (22%; n = 218), Iquique (14.3%; n = 142), Alto Hospicio (9.5%; n = 94), Antofagasta (37.1%; n = 368) and Calama (17.1%; n = 170). A total of 82.4% (n = 818) said they were heterosexual, and 56.8% (n = 565) said they had used protective barrier methods in the last 2 years. Sociodemographic details are shown in **Supplementary Table 1**.

Instruments

Condom Use Attitude Scale (CUAS): developed ad-hoc to evaluate the subjective valence of prevention behaviors and use of protective barriers through three attitudinal dimensions: (a) affective, (b) behavioral, and (c) cognitive. The response options are in a Likert format of 4 ordered categories (1 = "Strongly disagree" to 4 = "Strongly agree"). The statements refer to negative attitudes/behavior toward condom use. Therefore, high scores suggest an unfavorable attitude toward condom use.

Initially, 60 items (20 for each dimension) were outlined and assessed by three expert judges (two health professional judges and one judge with psychometric experience) based on grammatical adequacy (coherence and clarity) and construct representativeness. Judges individually scored "1, 0, or −1," where "1" represents grammatical adequacy and construct representativeness of the item. After that, means were calculated, and items with means less than or equal to 0 were eliminated. A pilot study was applied online (n = 110), with a 32-item version, from which those items with values below 0.30 in the corrected homogeneity coefficient were iteratively eliminated. Finally, a 17-item version was obtained, applied in samples (a) and (b) for this study. The final version (see **Supplementary Protocols 1, 2**) and its psychometric evidence are reported in the results section.

Sexual risk behaviors scale (Ferrer-Urbina et al., 2018): is a 12-item instrument designed to assess 3 dimensions of sexual risk behaviors: (a) sexual activity with multiple partners (4 items), (b) inappropriate or insufficient use of protective barriers (4 items), and (c) sexual activity under the influence of alcohol and drugs (4 items). The response options have a Likert format of four ordered categories (1 = “Strongly Disagree” to 4 = “Strongly Agree”) to avoid acquiescence and to force decision making given their attitudinal character. Response options are conditional on reporting only behaviors in the past 2 years. The scale stated evidence of validity based on internal structure and adequate levels of reliability ($\omega > 0.8$) (Ferrer-Urbina et al., 2018).

Procedure

Eight fifth-year psychology students were trained to apply the questionnaires in pencil and paper format to collect the samples (a). Participants were contacted between March and May 2018 in recreational areas of higher education institutions (e.g., reading areas, interior courtyards, library, and others.). The response procedure lasted approximately 15 min.

Twenty surveyors were trained and assigned in the study cities to collect the sample (b). Participants were contacted between March and July 2019 in the busiest areas of each city. The questionnaires were self-administered in pencil and paper format. The answer procedure lasted 15–20 min.

In both sample collection processes, the questionnaire was provided with an informed consent, where the research objectives, the rights of the participants, anonymity, and confidentiality of their participation were established. Volunteers responded on the spot without any reward or incentive. Anonymity was guaranteed by the return of the questionnaire in a sealed envelope, without any personal identification. The research was approved by the Scientific Ethics Committee of the Universidad de Tarapacá, within the framework of ANID’s grant (Fondecyt de iniciación 11170395).

Data Analysis

Initially, to establish the empirical dimensionality of the test with sample (a), a parallel analysis was realized, based on exploratory factor analysis (EFA) with minimum residual estimation method, 20 replicates, and based on average eigenvalues of the simulated data. Then, to establish the preliminary evidence of validity based on the internal structure of the test, an ESEM with GEOMIN rotation (Asparouhouv and Muthén, 2009) and WLSMV estimation method, which is robust with non-normal discrete variables (DiStefano and Morgan, 2014; Li, 2016), was performed from the polychoric correlation matrix, given the ordinal structure of the data (Barendse et al., 2015). Subsequently, in order to obtain a shorter and optimized scale, it was iteratively debugged by eliminating items based on three criteria: (1) retention of items with strong factor loadings ($\lambda > 0.5$); (2) elimination of redundant items (Abad et al., 2011) and (3) elimination of items with strong cross-loadings (>0.3) (Muthén and Asparouhov, 2012; Xiao et al., 2019).

To test the empirical dimensionality of the debugged version, establish the evidence of the internal structure, contrast alternative models, assess the stability of the scale, and obtain evidence of validity based on the relationship with other variables, sample (b) was used. The dimensionality of the debugged version was tested with parallel analysis. The debugged factor structure was tested with ESEM with GEOMIN rotation and an alternative one-dimensional model, both with the WLSMV estimation method and based on polychoric correlations. Reliability was also estimated for each dimension and with Cronbach’s alpha and McDonald’s omega coefficients, both in non-ordinal versions (Viladrich et al., 2017). Measurement invariance between persons of different sexes was assessed using a multi-group ESEM (i.e., metric and scalar). Increases in RMSEA below 0.010 were considered evidence of invariance (Chen, 2007). Finally, evidence of validity based on the relationship with other variables was established, using SET-ESEM (with GEOMIN rotation, WLSMV estimator and from polychoric correlations), between the dimensions of the CUAS, the dimensions of the sexual risk behavior scale (Ferrer-Urbina et al., 2018), and a single-item scale on condom use in the past 2 years.

The overall organization of the models was assessed following the cut-point recommendations stated by Schreiber (2017) for the comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean squared error of approximation (RMSEA) (e.g., CFI > 0.95; TLI > 0.95; RMSEA < 0.06).

Reliability and parallel analysis were performed with Jamovi version 2.0.0 (The Jamovi Project, 2020) and ESEM, SET-ESEM and Measurement invariance were performed with Mplus version 8.2 (Muthén and Muthén, 1998-2017).

Sample (a) showed 0.017% of missing data, while the sample (b) showed 0.027% of missing data. The Pairwise method was used for the missing data handling.

RESULTS

Exploratory Analysis With Sample (a)

Initially, parallel analysis suggested four factors with eigenvalues over the extraction of random variables, but one of them was explained only by two items. Then, excluding those two items, a second parallel analysis suggest three factors (see **Supplementary Figure 1**), showed good statistic’s fit ($\chi^2_{DF=63}$ = 161.437; TLI = 0.929; RMSEA = 0.056) and factorials loadings that fit with the proposed structure (see **Supplementary Table 2**). Then, two ESEM models were estimated, one with the original version (M1) (three dimensions and 15 items) and the other with the debugged version (M2a) (three dimensions and 10 items), using the reported criteria in data analysis section. The original model (M1) showed good statistic's fit, according to the standards recommended in the literature ($\chi^2_{DF=63}$ = 179.230; CFI = 0.982; TLI = 0.969; RMSEA = 0.060; SRMR = 0.031) (Schreiber, 2017). However, M2a ($\chi^2_{DF=18}$ = 61.566; CFI = 0.992; TLI = 0.979; RMSEA = 0.068; SRMR = 0.023) shown a better structure's fit, except for the RMSEA index (see **Supplementary Table 3**). The M2a model also showed good factor loadings on each factor (affective, λ = 0.55–0.91; behavioral, λ = 0.58–0.83; cognitive, λ = 0.47–0.79) and low levels of cross-loadings (affective, λ = −0.07–0.11; behavioral, λ = −0.04–0.22; cognitive, λ = −0.07–0.04). Structural relationships between dimensions were moderate ($r > 0.30$) and mild ($r > 0.10$) (Cohen, 1988; see **Supplementary Table 4**).

Dimensionality and Evidence of Validity Based on Internal Structure

In sample (b), from the debugged (10 items version), a parallel analysis, an ESEM (M2b) (three dimensions with 10 items) and one-dimensional (M3) models was tested. The parallel analysis suggested three factors with eigenvalues over the extraction of random variables (see **Supplementary Figure 2**). Only M2b ($\chi^2_{DF=18}$ = 71.996; CFI = 0.994; TLI = 0.984; RMSEA = 0.055; SRMR = 0.016) showed good fit. M3 ($\chi^2_{DF=40}$ = 1480.23; CFI = 0.829; TLI = 0.780; RMSEA = 0.205; SRMR = 0.115) showed fit indicators far away from the standards recommended in the literature. Details of the validity based on the internal structure of the test are shown in **Supplementary Table 3**.

Standardized Factor Loadings, Factorial Covariations and Score's Reliability

Table 1 presents the factor loadings with their corresponding factorial covariates and reliability coefficients of the three-dimensional covariate (M2b) and the one-dimensional (M3) models in the sample (b).

M2b model, factor loadings were shown to be adequate or at least sufficient representations of each factor (affective, λ = 0.60–0.92; behavioral, λ = 0.38–0.77; cognitive, λ = 0.67–0.83) and to have low levels of cross-loadings (affective, λ = –0.01–0.35; behavioral, λ = –0.01–0.11; cognitive, λ = –0.01–0.27). Structural relationships between dimensions were moderate ($r > 0.30$) (Cohen, 1988). Reliability estimates were adequate or at least sufficient (ω > 0.70; α > 0.70) (Cho and Kim, 2015), although slightly lower (ω > 0.60; α > 0.60), in the case of the behavioral dimension.

In M3 model, factor loadings were shown to be adequate representations of one-dimensional factor (λ = 0.49 –0.85). Details of standardized factor loadings, factorial covariations and score's reliability are shown in **Table 1**.

Factorial Invariance and Evidence of Validity Based on the Relationship With Other Variables

The three-dimensional covariate model (M2b) was used to estimate tests of invariance between men and women, with sample (b). The metric (RMSEA = 0.049) and scalar model (RMSEA = 0.054) (restricted) compared to the configural model (RMSE = 0.049) (unrestricted) showed no relevant changes in the RMSEA differential, with the equivalence between factor loadings and factor thresholds being sustainable for both groups. Details of factorial invariance test are shown in **Supplementary Table 3**.

Finally, the SET-ESEM model that estimated the association between the latent dimensions of the attitudes regarding condom use scale and the sexual risk behaviors scale, in addition to the observed variable condom use (see **Figure 1**), showed satisfactory comparative and absolute fit indexes ($\chi^2_{DF=181}$ = 395.128; CFI = 0.986; TLI = 0.980; RMSEA = 0.035, [CI = 0.030–0.039]). According to observed relationships, the affective component had mild direct effects on sexual activity with multiple partners (γ = 0.216, $p < 0.001$); sexual activity under the influence of alcohol or drugs (γ = 0.237, $p < 0.001$); and a slight effect on condom use (γ = 0.141, p = 0.012). The behavioral component had a large effect on inappropriate use of protective barriers (γ = 0.736, $p < 0.001$) and a moderate inverse effect on condom use (γ = –0.452, $p < 0.001$). Finally, the cognitive component had a moderate inverse effect on inappropriate use of protective barriers (γ = –0.455, $p < 0.001$). Details of standardized relationships between latent dimensions are shown in **Supplementary Table 5**.

DISCUSSION

The purpose of this study was to develop a scale to measure attitudes toward condom use, with evidence of validity for its use in youth and young adults in a Chilean context. The fit statistics of the covariate model (M2), both for the sample of students (a) and the general population (b), the size of the factor loadings, and the absence of relevant cross-loadings (except for one item of the behavioral dimension that showed a slightly higher cross-loading with the affective dimension), support the multidimensional structure of the model (three dimensions) and provide evidence of validity based on the internal structure for the adequate interpretation of the scores. The one-dimensional model (M3) cannot be used as a plausible explanation of the instrument so that the scores could be interpreted from its specific aspects and do not from the combination of its dimensions. The estimates of the reliability coefficients allow us to argue that each dimension has an adequate or satisfactory level of consistency, except for the behavioral dimension, which has a slightly lower estimate, which is to be expected given the small number of items and the effect this has on reliability estimates based on internal consistency.

According to the invariance standards suggested by Chen (2007), it is possible to sustain metric and scalar invariance of the measures according to the gender of the participants (i.e.,

TABLE 1 | Descriptive information of the CUAS and factor loadings resulting from ESEM and CFA in general population sample.

Attitude toward condoms	Mean (*SD*)	S	K	Factor loadings				Reliability	
				M2b			M3	α if item is dropped	ω if item is dropped
				A	B	C	–		
Affective (A)									
Es difícil disfrutar del sexo cuando se usa preservativo. (It's hard to enjoy sex when you use a condom).	2.44 (0.98)	–0.57	6.55	**0.830****	0.080	–0.018	**0.819****	0.731	0.731
Siento que el preservativo disminuye mi satisfacción sexual. (I feel that the condom decreases my sexual satisfaction).	2.51 (0.97)	–2.28	6.22	**0.924****	0.011	–0.010	**0.854****	0.681	0.681
Las personas obtienen más placer en las relaciones sexuales sin preservativo. (People get more pleasure from sex without a condom).	2.63 (1.06)	–3.06	–7.54	**0.608****	–0.014	0.271**	**0.679****	0.849	0.848
Behavioral (B)									
Evito usar preservativo cada vez que me lo permiten. (I avoid using a condom every time I'm allowed).	2.17 (1.03)	4.60	–6.82	0.356**	**0.498****	0.005	**0.622****	0.504	0.504
No suelo llevar preservativos cuando tengo un encuentro sexual. (I don't usually wear a condom when I have a sexual encounter).	2.26 (1.09)	3.40	–8.03	–0.012	**0.776****	–0.017	**0.494****	0.503	0.502
Tendría relaciones sexuales aun cuando mi pareja se negará a usar preservativo. (I would have sex even if my partner refused to use a condom).	2.33 (1.05)	1.24	–783	0.174*	**0.381****	0.183**	**0.525****	0.588	0.588
Cognitive (C)									
Creo que el preservativo debieran usarlo solo las personas promiscuas. (I think the condom should only be used by promiscuous people).	1.63 (0.93)	17.54	5.38	–0.004	0.113	**0.714****	**0.670****	0.725	0.729
El uso de preservativos es solo para relaciones pasajeras. (The use of condoms is only for temporary relations).	1.84 (0.99)	11.05	–3.08	–0.003	0.084	**0.775****	**0.707****	0.690	0.695
Pienso que el preservativo es innecesario en las personas sanas. (I think condoms are unnecessary in healthy people).	1.65 (0.89)	15.53	3.34	–0.005	–0.003	**0.838****	**0.705****	0.687	0.699
Creo que sugerir el uso del preservativo genera desconfianza. (I think that suggesting condom use creates distrust).	1.69 (0.90)	14.58	1.87	0.079	–0.009	**0.673****	**0.598****	0.754	0.756
					Correlations		–	α index	ω index
Affective (A)	2.53 (0.86)	–2.26	–5.79	–			–	0.822	0.826
Behavioral (B)	2.25 (0.80)	0.32	–3.79	**0.445****	–		–	0.631	0.635
Cognitive (C)	1.69 (0.71)	1.16	1.34	**0.318****	**0.459****	–	–	0.770	0.774

*SD, Standard Deviation; S, Skewness; K, Kurtosis. **$p < 0.001$; M2b, ESEM with three covariate factors; 10 items, M3; One-dimensional CFA, 10 items.*
Bold values indicate the expected dimension of the factor loadings.
$^*p < 0.05$.

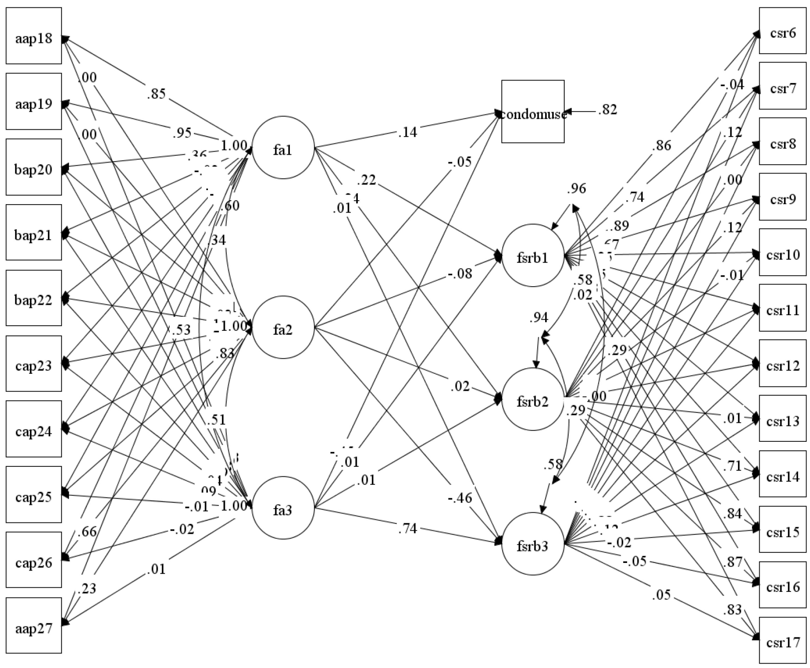

FIGURE 1 | Graphical representation of the SET-SEM model; Fa1, Affective; Fa2, Cognitive; Fa3, Behavioral; Fsrb1, Sexual activity with multiple partners; Fsrb2, Sexual activity under the influence of alcohol or drugs; Fsrb3, Inappropriate use of protective barriers; condomuse, Condom use.

strong invariance). Therefore, it is possible to apply the scale to men and women and compare them since the factor loadings were equivalent between the two groups and the dimensions had the same differential variability between the sexes. Although it is crucial to consider the effects produced by socio-structural variables such as gender roles on condom use (Casique, 2019), this scale allows comparisons between men and women with an adequate interpretation of the results obtained.

In terms of evidence of validity based on association with other variables, the dimensions of the CUAS were observed to explain sexual risk behaviors and condom use partially, as had been noted in previous research (Smith et al., 2015; Johnson et al., 2018; Kilwein and Looby, 2018; Mbelle et al., 2018; Yi et al., 2018; Ajayi et al., 2019). The observed relationships were in the expected direction, except for the relationship between the affective component with condom use. This exception could be attributed to the involvement of other variables of relevance to condom use (i.e., risk perception, knowledge of risky sexual behaviors, communication with sexual partners) that overlap with people's affective attitudes toward condom use. Therefore, it is possible that people who manifest a negative attitude might not necessarily reduce condom use, especially if they believe they are susceptible to the risk of transmission or are aware of the risks associated with non-use (Sheeran et al., 2016; Janulis et al., 2017; Teye-Kwadjo et al., 2017; Elshiekh et al., 2020).

The main constraint of this study corresponds to the size and representativeness of the sample. Although two samples were used, both were non-probabilistic. Therefore, there is no guarantee of the adequacy of the generalization to population values. It is suggested that further psychometric studies be conducted using this instrument in other populations (e.g., high-risk populations, new countries, and migrant populations) and medical, health, and educational contexts. Finally, we need to be careful about the relation of the self-report item of condom use, since the extremely large effect size can be explained, in part, by the fact that the behavioral dimension of the scale makes direct references to condom use behavior.

The inclusion of this scale in an evaluation protocol in health services or educational establishments could be helpful since the information provided by this measurement instrument will make it possible to identify groups of subjects that require specific preventive interventions. Consequently, current interventions and strategies that promote sexual health among youth and adults could be complemented and improved.

The final version (10 items) of the scale of attitudes toward condom use presents evidence of reliability and validity (i.e., based on the internal structure of the test and the association with other variables). The initial evidence suggests that the current scale constitutes a new brief instrument developed with contemporary psychometric techniques and establishes an updated and alternative proposal to assess attitudes toward condom use, which can also be used to develop studies on the psychological factors involved in sexual behaviors.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee of the Universidad de Tarapacá. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RF-U: conceptualization, formal analysis, methodology, writing—original draft, and writing—review and editing. PM-C and GS-P: methodology, data curation, data analysis, writing—original draft, and writing—review and editing. MC-H: formal analysis and writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by ANID (National Agency for Research and Development), through Grant FONDECYT de Iniciación N° 11170395.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.727499/full#supplementary-material

REFERENCES

Abad, F., Olea, J., Ponsoda, V., and García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Madrid: Síntesis.

Ajayi, A., Ismail, K., and Akpan, W. (2019). Factors associated with consistent condom use: a cross-sectional survey of two Nigerian universities. *BMC Public Health* 19:1207. doi: 10.1186/s12889-019-7543-1

Ajzen, I., and Madden, T. J. (1986). Prediction of goal-directed behavior: attitudes, intentions, and perceived behavioral control. *J. Exp. Soc. Psychol.* 22, 453–474.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Asparouhouv, T., and Muthén, B. (2009). Exploratory structural equation modeling. *Struct. Equ. Modeling* 16, 397–438. doi: 10.1080/10705510903008204

Ato, M., López-García, J., and Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anal. Psicol.* 29, 1038–1059. doi: 10.6018/analesps.29.3.178511

Barendse, M., Oort, F., and Timmerman, M. (2015). Using exploratory factor analysis to determine the dimensionality of discrete responses. *Struct. Equ. Modeling* 22, 87–101. doi: 10.1080/10705511.2014.934850

Barrera-Herrera, A., and Vinet, E. (2017). Adultez emergente y características culturales de la etapa en universitarios chilenos. *Ter. Psicol.* 35, 47–56. doi: 10.4067/S0718-48082017000100005

Beachy, S., Lechuga, J., Dickson-Gomez, J., and Liang, C. (2020). Validation of brief condom use attitudes scales for spanish speaking people-who-use-drugs in El Salvador. *Research Square* [Preprint] doi: 10.21203/rs.3.rs-20069/v1

Brown, I. (1984). Development of a scale to measure attitude toward the condom as a method of birth control. *J. Sex Res.* 20, 255–263. doi: 10.1080/00224498409551224

Bryan, A., Norris, J., Abdallah, D., Zawacki, T., Morrison, D., George, W., et al. (2017). Condom-insistence conflict in Women's alcohol-involved sexual encounters with a new male partner. *Psychol. Women Q.* 41, 100–113. doi: 10.1177/0361684316668301

Cáceres, K., and Pino, R. (2018). Estimaciones poblacionales sobre VIH en Chile 2017 SPECTRUM ONUSIDA. *Rev. Chil. Infectol.* 35, 642–648. doi: 10.4067/S071610182018000600642

Cáceres-Burton, K. (2019). Informe: situación epidemiológica de las infecciones de transmisión sexual en Chile, 2017. *Rev. Chil. Infectol*. 36, 221–233. doi: 10.4067/S0716-10182019000200221

Casique, I. (2019). Gender differences in the sexual well-being of Mexican adolescents. *Int. J. Sex Health* 31, 1–16. doi: 10.1080/19317611.2018.1561587

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Modeling* 14, 464–504. doi: 10.1080/10705510701301834

Cho, E., and Kim, S. (2015). Cronbach's coefficient alpha: well known but poorly understood. *Organ. Res. Methods* 18, 207–230. doi: 10.1177/1094428114555994

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*, 2nd Edn. New York, NY: Routledge.

Crosby, R., Graham, C., Milhausen, R., Sanders, S., and Yarber, W. (2010). "Condom use errors/problems survey," in *Handbook of Sexuality-Related Measures*, eds T. D. Fisher, C. Davis, W. Yarber, and S. Davis (New York, NY: Routledge), 153–159. doi: 10.1097/OLQ.0000000000000356

Davis, K. C., Kirwan, M., Wegner, R., Neilson, E. C., and Stappenbeck, C. A. (2020). Effects of alcohol, condom request style, and state anger on men's condom use resistance. *J. Stud. Alcohol Drugs* 81, 454–461. doi: 10.15288/jsad.2020.81.454

De Torres, R. (2020). Facilitators and barriers to condom use among Filipinos: a systematic review of literature. *Health Promot. Perspect.* 10:306. doi: 10.34172/hpp.2020.49

DiStefano, C., and Morgan, G. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Struct. Equ. Modeling* 21, 425–438. doi: 10.1080/10705511.2014.915373

Elshiekh, H., Hoving, C., and de Vries, H. (2020). Exploring determinants of condom use among university students in Sudan. *Arch. Sex. Behav.* 49, 1379–1391. doi: 10.1007/s10508-019-01564-2

Ferrer-Urbina, R., Leal-Soto, F., Bravo, N., Huaranca, C., Perez, J., Salinas, T., et al. (2018). Scale of risk behaviors, associated with STI / HIV-AIDS, for young Chileans. *Eur. Proc. Soc. Behav. Sci.* 60, 800–809. doi: 10.15405/epsbs.2019.04.02.99

Fishbein, M., and Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

Gaborieau, J. B., and Pronello, C. (2021). Validation of a unidimensional and probabilistic measurement scale for pro-environmental behaviour by travellers. *Transportation* 48, 555–593. doi: 10.1007/s11116-019-10068-w

Habel, M., Leichliter, J., Dittus, P., Spicknall, I., and Aral, S. (2018). Heterosexual anal and oral sex in adolescents and adults in the United States, 2011–2015. *Sex. Transm. Dis.* 45:775. doi: 10.1097/OLQ.0000000000000889

Hanna, K. (1999). An adolescent and young adult condom self-efficacy scale. *J. Pediatr. Nurs.* 14, 59–66. doi: 10.1016/S0882-5963(99)80061-X

Helweg-Larsen, M., and Collins, B. (1994). The UCLA multidimensional condom attitudes scale: documenting the complex determinants of condom use in college students. *Health Psychol.* 13, 224–237. doi: 10.1037/0278-6133.13.3.224

Hollub, A., Reece, M., Herbenick, D., Hensel, D., and Middlestadt, S. (2011). College students and condom attitude: validation of the multi-factor attitude toward condoms scale (MFACS). *J. Am. Coll. Health* 59, 708–714. doi: 10.1080/07448481.2010.546462

Instituto de Salud Pública (2016). *Boletín Vigilancia de Laboratorio. Resultados Confirmación de Infecciones por VIH en Chile, 2010–2015*. Available online at: https://www.ispch.cl/sites/default/files/BoletinVIH-15112017A.pdf (accessed May 04, 2021)

Instituto Nacional de Estadística (2018). *Síntesis de Resultados CENSO 2017*. Available online at: https://www.censo2017.cl/descargas/home/sintesis-de-resultados-censo2017.pdf (accessed May 04, 2021)

Janulis, P., Newcomb, M., Sullivan, P., and Mustanski, B. (2017). Evaluating HIV knowledge questionnaires among men who have sex with men: a multi-study item response theory analysis. *Arch. Sex. Behav.* 47, 107–119. doi: 10.1007/s10508-016-0910-4

Javier, S., Abrams, J., Moore, M., and Belgrave, F. (2018). Condom use efficacy and sexual communication skills among African American college women. *Health Promot. Pract.* 19, 287–294. doi: 10.1177/1524839916676253

Johnson, W., O'Leary, A., and Flores, S. (2018). Per-partner condom effectiveness against HIV for men who have sex with men. *AIDS* 32, 1499–1505. doi: 10.1097/QAD.0000000000001832

Joint United Nations Program on HIV/AIDS (2019). *UNAIDS Data 2019.* Available online at: https://www.unaids.org/en/resources/documents/2019/2019-UNAIDS-data (accessed May 04, 2021)

Kilwein, T., and Looby, A. (2018). Predicting risky sexual behaviors among college student drinkers as a function of event-level drinking motives and alcohol use. *Addict. Behav.* 76, 100–105. doi: 10.1016/j.addbeh.2017.07.032

Kilwein, T., Kern, S., and Looby, A. (2017). Interventions for alcohol-related risky sexual behaviors among college students: a systematic review. *Psychol. Addict. Behav.* 31, 944–950. doi: 10.1037/adb0000294

Kim, Y., Min, H., Lee, J., and Kim, S. (2021). Una revisión integradora de estudios sobre el uso de condones entre estudiantes universitarios coreanos. *Invest. Enferm. Salud Infantil.* 27, 43–55. doi: 10.4094/chnr.2021.27.1.43

La Trobe, H. L., and Acott, T. G. (2000). A modified NEP/DSP environmental attitudes scale. *J. Environ. Educ.* 32, 12–20. doi: 10.1080/00958960009598667

Li, C. (2016). Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* 48, 936–949. doi: 10.3758/s13428-015-0619-7

Mbelle, N., Mabaso, M., Chauke, T., Sigida, S., Naidoo, D., and Sifunda, S. (2018). Perception and attitudes about male and female condom use amongst university and technical and vocational education and training (TVET) college students in South Africa: a qualitative enquiry of the 2014 higher education and training HIV/AIDS (HEAIDS) programme first things first campaign. *J. HIV AIDS* 4, 1–8.

Mileti, F., Mellini, L., Sulstarova, B., Villani, M., and Singy, P. (2018). Exploring barriers to consistent condom use among sub-Saharan African young immigrants in Switzerland. *AIDS Care* 31, 113–116. doi: 10.1080/09540121.2018.1526371

Ministerio de Salud (2017). *ENCUESTA NACIONAL DE SALUD 2016-2017, Primeros Resultados*. Available online at: https://www.minsal.cl/wp-content/uploads/2017/11/ENS-2016-17_PRIMEROS-RESULTSADOS.pdf (accessed May 04, 2021)

Muthén, B., and Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychol. Methods* 17:313. doi: 10.1037/a0026802

Muthén, L., and Muthén, B. (1998-2017). *Mplus User's Guide*, 8th Edn. Los Angeles, CA: Muthén & Muthén.

Neilands, T., and Choi, K. (2002). A validation and reduced form of the female condom attitudes scale. *AIDS Educ. Prev.* 14, 158–171. doi: 10.1521/aeap.14.2.158.23903

Otzen, T., and Manterola, C. (2017). Técnicas de Muestreo sobre una Población a Estudio. *Int. J. Morphol.* 35, 227–232. doi: 10.4067/S0717-95022017000100037

Pan American Health Organization (2019). *Situación de la Epidemia de la Infección por el VIH y Respuesta, América Latina y el Caribe, 2019.* Available online at: https://www.paho.org/hq/index.php?option=com_docman&view=download&category_slug=datos-estadisticos-5691&alias=51070-situacion-de-la-epidemia-de-la-infeccion-por-el-vih-y-respuesta-america-latina-y-el-caribe-2019&Itemid=270&lang=es (accessed May 04, 2021)

Patterson, Y. (2019). Intragroup differences among African jamaican and African American women: empowerment, male condom-use intentions and negotiation. *Soc. Work Public Health* 34, 1–19. doi: 10.1080/19371918.2019.1589612

Pinyaphong, J., Srithanaviboonchai, K., Chariyalertsak, S., Phornphibul, P., Tangmunkongvorakul, A., and Musumari, P. (2018). Inconsistent condom use among male university students in northern Thailand. *Asia Pac. J. Public Health* 30, 147–157. doi: 10.1177/1010539517753931

Plaza-Vidal, R., Ibagon-Parra, M., and Vallejo-Medina, P. (2020). Spanish translation, adaptation, and validation of the multidimensional condom attitudes scale with young Colombian men and women. *Arch. Sex. Behav.* 50, 2729–2740. doi: 10.1007/s10508-020-01759-y

Sacco, W., Levine, B., Reed, D., and Thompson, K. (1991). Attitudes about condom use as an AIDS-relevant behavior: their factor structure and relation to condom use. *J. Consult. Clin. Psychol.* 3, 265–272. doi: 10.1037/1040-3590.3.2.265

Schreiber, J. B. (2017). Update to core reporting practices in structural equation modeling. *Res. Soc. Adm. Pharm.* 13, 634–643. doi: 10.1016/j.sapharm.2016.06.006

Shamu, S., Khupakonke, S., Farirai, T., Slabbert, J., Chidarikire, T., Guloba, G., et al. (2020). Knowledge, attitudes and practices of young adults towards HIV prevention: an analysis of baseline data from a community-based HIV prevention intervention study in two high HIV burden districts, South Africa. *BMC Public Health* 20:1249. doi: 10.1186/s12889-020-09356-3

Sheeran, P., Abraham, C., and Orbell, S. (1999). Psychosocial correlates of heterosexual condom use: a meta-analysis. *Psychol. Bull.* 125, 90–132. doi: 10.1037/0033-2909.125.1.90

Sheeran, P., Maki, A., Montanaro, E., Avishai-Yitshak, A., Bryan, A., Klein, W. M., et al. (2016). The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: a meta-analysis. *Health Psychol.* 35, 1178–1188. doi: 10.1037/hea0000387

Smith, D., Herbst, J., Zhang, X., and Rose, C. E. (2015). Condom effectiveness for HIV prevention by consistency of use among men who have sex with men in the United States. *J. Acquir. Immune Defic. Syndr.* 68, 337–344. doi: 10.1097/QAI.0000000000000461

Teye-Kwadjo, E., Kagee, A., and Swart, H. (2017). Predicting the intention to use condoms and actual condom use behaviour: a three-wave longitudinal study in Ghana. *Appl. Psychol. Health Well Being* 9, 81–105. doi: 10.1111/aphw.12082

The Jamovi Project (2020). *Jamovi (Version 1.8.1) [Computer Software]*. Available online at: https://www.jamovi.org (accessed 26 March 2021)

Vallejo-Medina, P., Ramírez, E., Saavedra-Roa, A., Gómez-Lugo, M., and Pérez-Durán, C. (2019). Spanish validation of female condom attitude scale and female condom use in Colombian young women. *BMC Women's Health* 19:128. doi: 10.1186/s12905-019-0825-z

Viladrich, C., Angulo-Brunet, A., and Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Anal. Psicol.* 33, 755–782. doi: 10.6018/analesps.33.3.268401

World Health Organization (2016). *Proyecto de Estrategia Mundial del Sector de la Salud Contra las Infecciones de Transmisión Sexual Para 2016-2021.* Available online at: http://www.who.int/reproductivehealth/GHSS_STI_SP_06012016.pdf (accessed May 04, 2021)

Xiao, Y., Liu, H., and Hau, K. (2019). A comparison of CFA, ESEM, and BSEM in test structure analysis. *Struct. Equ. Modeling* 26, 665–677. doi: 10.1080/10705511.2018.1562928

Yi, S., Te, V., Pengpid, S., and Peltzer, K. (2018). Social and behavioural factors associated with risky sexual behaviours among university students in nine ASEAN countries: a multi-country cross-sectional study. *SAHARA J.* 15, 71–79. doi: 10.1080/17290376.2018.1503967

Zanna, M. P., and Rempel, J. K. (1988). "Attitudes: a new look at an old concept," in *The Social Psychology of Knowledge*, eds D. Bar-Tal and A. W. Kruglanski (Cambridge: Cambridge University Press), 315–334.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 07 December 2021
doi: 10.3389/fpsyg.2021.754103

Check for updates

Are the Items of the Starkstein Apathy Scale Fit for the Purpose of Measuring Apathy Post-stroke?

Stanley Hum[1,2]*, Lesley K. Fellows[1,2], Christiane Lourenco[3] and Nancy E. Mayo[1,4,5]

[1] McGill University, Montreal, QC, Canada, [2] Montreal Neurological Institute, Montreal, QC, Canada, [3] Escola Superior de Ciências da Santa Casa de Misericórdia de Vitória, Vitória, Brazil, [4] McGill University Health Centre (MUHC), Montreal, QC, Canada, [5] McGill University Health Centre-Research Institute (MUHC-RI), Montreal, QC, Canada

OPEN ACCESS

Edited by:
Victor Zaia,
Faculdade de Medicina do ABC,
Brazil

Reviewed by:
Purya Baghaei,
Islamic Azad University of Mashhad,
Iran
Theodoros Kyriazos,
Panteion University, Greece

Correspondence:
Stanley Hum
Stanley.hum@mcgill.ca

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *05 August 2021*
Accepted: *25 October 2021*
Published: *07 December 2021*

Citation:
Hum S, Fellows LK, Lourenco C and Mayo NE (2021) Are the Items of the Starkstein Apathy Scale Fit for the Purpose of Measuring Apathy Post-stroke? Front. Psychol. 12:754103. doi: 10.3389/fpsyg.2021.754103

Importance: Given the importance of apathy for stroke, we felt it was time to scrutinize the psychometric properties of the commonly used Starkstein Apathy Scale (SAS) for this purpose.

Objectives: The objectives were to: (i) estimate the extent to which the SAS items fit a hierarchical continuum of the Rasch Model; and (ii) estimate the strength of the relationships between the Rasch analyzed SAS and converging constructs related to stroke outcomes.

Methods: Data was from a clinical trial of a community-based intervention targeting participation. A total of 857 SAS questionnaires were completed by 238 people with stroke from up to 5 time points. SAS has 14 items, rated on a 4-point scale with higher values indicating more apathy. Psychometric properties were tested using Rasch partial-credit model, correlation, and regression. Items were rescored so higher scores are interpreted as lower apathy levels.

Results: Rasch analysis indicated that the response options were disordered for 8/14 items, pointing to unreliability in the interpretation of the response options; they were consequently reduced from 4 to 3. Only 9/14 items fit the Rasch model and therefore suitable for creating a total score. The new rSAS was deemed unidimensional (residual correlations: < 0.3), reasonably reliable (person separation index: 0.74), with item-locations uniform across time, age, sex, and education. However, 30% of scores were > 2 SD above the standardized mean but only 2/9 items covered this range (construct mistargeting). Apathy (rSAS/SAS) was correlated weakly with anxiety/depression and uncorrelated with physical capacity. Regression showed that the effect of apathy on participation and health perception was similar for rSAS/SAS versions: R^2 participation measures ranged from 0.11 to 0.29; R^2 for health perception was ~0.25. When placed on the same scale (0–42), rSAS value was 6.5 units lower than SAS value with minimal floor/ceiling effects. Estimated change over time was identical

(0.12 units/month) which was not substantial (1.44 units/year) but greater than expected assuming no change (t: 3.6 and 2.4).

Conclusion: The retained items of the rSAS targeted domains of behaviors more than beliefs and results support the rSAS as a robust measure of apathy in people with chronic stroke.

Keywords: Rasch analysis, stroke, apathy, measurement, patient-reported outcome, modern psychometrics

INTRODUCTION

Apathy is a defining feature in many common neurological conditions, including Parkinson's Disease, Alzheimer's Disease, and stroke (Robert et al., 2018; Le Heron et al., 2019). A meta-analysis of 24 studies found that apathy occurred in 30–40% of stroke patients (van Dalen et al., 2013). Mayo et al. (2009) estimated from an inception cohort that 20% of patients had apathy, as reported by a close companion, at some point in the first year post-stroke. They also found that apathy strongly affected recovery. Apathy has also been shown to affect health related quality of life (HRQL) in patients with stroke (Tang et al., 2014).

Although apathy is recognized as common and clinically important, there have been challenges in defining and measuring this construct. Marin originally described apathy as the "lack of motivation not attributable to disturbance of intellect, emotion, or level of consciousness" (Marin, 1991) and operationalized the definition as "a state characterized by simultaneous diminution in the overt behavioral, cognitive, and emotional concomitants of goal-directed behavior" (Marin et al., 1991). Medical diagnostic criteria were developed for apathy in 2009 and revised in 2018. The 2018 consensus group largely echoed Marin's description, defining apathy as a quantitative reduction of goal-directed activity either in behavioral, cognitive, emotional, or social dimensions in comparison to the patient's previous level of functioning in these areas. They also indicated that these changes may be reported by patients themselves or be based on the observations of others.

Practically, apathy can be considered as a diagnosis (i.e., a binary outcome) and as a state that can be measured along a continuum. Measures that include a series of questions (items) or structured interview components about apathy behaviors, are typically used for both diagnosis and measurement.

Such questionnaires assess apathy from the perspectives of: (i) the patient using a patient-reported outcome (PRO) or a self-reported outcome (SRO); (ii) an observer who is usually a significant other (ObsRO); or (iii) a health care professional using a clinician reported outcome (ClinRO) (Mayo et al., 2017). There is an important distinction between PRO and SRO. The answer to the questions in a PRO can only be provided by the person without interpretation from anyone else (U.S. Department of Health and Human Services Food and Drug Administration, 2009). In SRO, the response given by the person can be amended based on other information that may not have been provided by the patient (Mayo et al., 2017).

A continuous measurement scale is constructed by assigning numerical labels to the ordinal item-responses (e.g., not at all, slightly, some, and a lot) and summing these labels, assuming they have mathematical properties, which they may not. The continuous scale can be used to quantify severity and change over time. A specific cut-point on this continuum is used for diagnostic classification.

Notwithstanding the consensus on defining apathy, there is a plethora of apathy questionnaires and no gold standard (Clarke et al., 2011). For example, a recent systematic review by Carrozzino (2019) identified 13 different apathy measures used in Parkinson's Disease research. Most of these have also been used in other neurological conditions, including stroke, without the due diligence required to ensure that the interpretation of the values apply to these other populations (Carrozzino, 2019).

As for many neuropsychiatric constructs, existing questionnaires were generally developed based on expert clinical input without a strong emphasis on psychometrics or patient experience to inform content. Classical approaches to psychometrics were applied after the items and the response options were set. The focus was on statistical homogeneity of the items (internal consistency expressed by Cronbach's alpha) and factor structure. Dimensionality of these measures was felt necessary to cover the construct, but not always reflected in the scoring. Further the psychometrics underlying the total score were not scrutinized; measures were formed as a simple sum of the ordinally labeled response options. These approaches have well-known drawbacks: A high internal consistency can arise from redundant items, and change on one item will yield similar changes in the redundant items resulting in an overestimate of the calculated amount of change. These legacy measures are now being scrutinized in light of modern psychometric methods such as Item Response Theory (IRT) and Rasch Measurement Theory (RMT) (Hobart and Cano, 2009; Petrillo et al., 2015).

RMT estimates the extent to which a set of items fits an underlying linear hierarchy (Rasch Model); items that do not fit this hypothesized model should not be included in the total score until revised (Mayo, 2015). Thus, psychometrics come into play before the items and response options are set. RMT tests whether response options are functioning as expected in terms of representing more (or less) of the construct and if not, modifications are made. Rasch analysis is also used to transform the ordinal scores of the response options to have mathematical interval-like properties, rather than just numerical labels, allowing a legitimate total score to be derived. This approach also allows testing whether the items reflect more than one dimension and, therefore, whether the construct is best reflected by multiple measures.

Another key feature of these modern psychometric methods is the ability to test whether the items have the same mathematical interpretation in different subpopulations, such as those defined by sex or gender, language, or by different disease groups (Pallant and Tennant, 2007; Tennant and Conaghan, 2007). The additional information on how the items respond is useful in creating a measure with strong properties, needed particularly for evaluating change over time.

The most commonly used self-reported questionnaire for assessing apathy in stroke are the 18 item Apathy Evaluation Scale (AES) developed by Marin et al. (1991) in 1991 and the Starkstein Apathy Scale (SAS) (Starkstein et al., 1992) which is based on a preliminary version of the AES. The SAS consists of a combination of 14 (PRO/SRO) items with a four-point ordinal scale (0 = not at all; 1 = slightly; 2 = some; and 3 = a lot), with a higher summed total score indicating more apathy. By summing the 14 items to generate a total score the original developer conceptualized the SAS as measuring a single apathy construct.

We were unable to find any studies of the psychometric properties of the SAS in stroke, although it has been used in that population (Starkstein et al., 1993). Given the importance of apathy for stroke, we felt it was time to scrutinize and, if needed, improve the SAS using modern psychometric approaches. The global aim of this study was to contribute evidence supporting the use of the SAS in people with stroke. The specific objectives were to: (i) to estimate the extent to which the items of the SAS fit a hierarchical continuum based on the Rasch Model; and (ii) to estimate the strength of the relationship between the Rasch analyzed SAS and converging constructs related to stroke outcomes.

MATERIALS AND METHODS

This study is a secondary analysis of an existing dataset from a clinical trial assessing a community-based, structured, program targeting participation in individuals post-stroke (Mayo et al., 2015). A sample of 238 English speaking participants living in the community having experienced a stroke within 5 years responded to the SAS at study entry (baseline = 0 month) and up to 4 more times (at 3, 6, 12, and > 12 months). A total of 857 SAS questionnaires were completed by the cohort over these time points and used in the Rasch analysis.

Measurement

The World Health Organization's (WHO) international classification for functioning, disability, and health (ICF) model) was used as the measurement framework and to structure the analyses as shown in **Figure 1**. Apathy is not part of the ICF, but motivation is listed as an impairment (b1301: mental functions that produce the incentive to act; the conscious or unconscious driving force for action). Apathy is part of the WHO international classification of diseases (ICD-10), listed as a diagnostic category not as a function (R45 symptoms and signs involving emotional state, specifically R45.3, demoralization and apathy). Apathy was represented by the original SAS with 14 items measured on a 4-point severity scale.

The literature supports a relationship between apathy and depression (both diagnostic categories; depression is ICD-10 F32-F34). Here, we included two measures of depressive symptoms (ICF b1265: impairment of optimism as defined by mental functions that produce a personal disposition that is cheerful, buoyant and hopeful, as contrasted to being downhearted, gloomy and despairing): Stroke-Specific Geriatric Depression Scale (SS-GDS) (Cinamon et al., 2011) and the anxiety/depression item of the Euroqol EQ-5D (EuroQol Group, 2016).

There is a very strong effect of impairment of apathy/motivation on physical function, participation and self-rated health (Mayo et al., 2014). Physical function was assessed here with measured gait speed (comfortable and maximum), hypothesized to be less affected by motivation as it is a measure of capacity to walk a short distance (5 m). Two measures of participation were available: Community Healthy Activities Model Program for Seniors (CHAMPS) (Stewart et al., 2001) using hours of meaningful activity in and outside the home as the metric and the Reintegration to Normal Living Index (RNL) (Wood-Dauphinee and Williams, 1987). The EQ-5D VAS (visual analog scale: 0—death to 100—perfect health) was the measure of health perception. Personal factors under consideration were gender, time (0; 3; 6; 12; > 12 months), age (< 50; 50–60; 61–70; 71–80; > 80 years), and education (≤ 12 years; > 12 years).

Statistical Methods

Rasch analysis was used to test whether the items on the SAS fit the Rasch model (Rasch, 1960). The typical steps and requirements for Rasch analysis were adopted from published guidelines (Pallant and Tennant, 2007; Tennant and Conaghan, 2007) and are detailed in **Figures 2**, **3**. RUMM2020 version 4.1 software was used. The process is iterative and items that do not fit are removed one at a time until all items fit the model.

The Likelihood ratio test statistic performed in RUMM2020 ($p < 0.0000001$) supports the partial credit model. The threshold distances varied across items supporting the use of a partial credit model for this analysis (see **Supplementary Appendix Figure 1** of item threshold map in **Supplementary Appendix**; Pallant et al., 2006).

Rasch analysis was conducted on all 14 items, initially. The number of observations available for Rasch analysis was 857 arising from 238 people with stroke, more than adequate to estimate item and person difficulties with a high degree of confidence level (99%) within a precision of ± 0.5 logits (Linacre, 1994). Please note that although 857 data points were entered into RUMM2020. The usable sample size was 856 since one data point had all extreme scores and was automatically removed. Bonferroni correction and/or *post hoc* downward sample size adjustment (features in RUMM2020) were used as appropriate (Hagell and Westergren, 2016; Hansen and Kjaersgaard, 2020). Five random samples with N of 300 each (with replacement) were drawn based on at least 10 observations per category (Linacre, 1999).

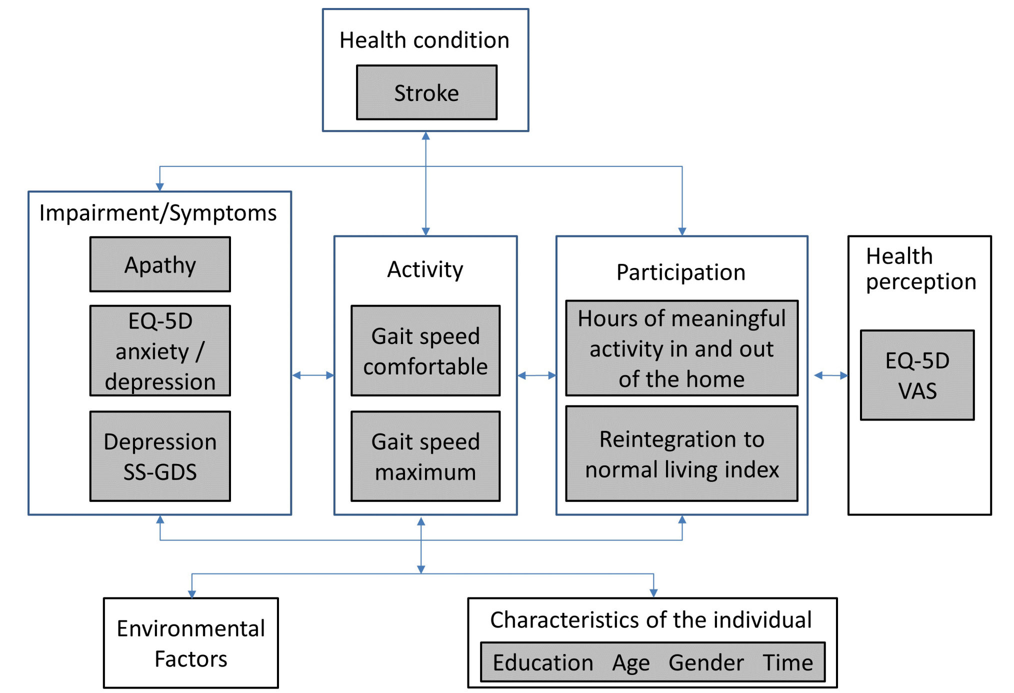

FIGURE 1 | WHO international classification of functioning, disability and health (ICF) for the measurement framework.

To access the extent to which using multiple time points introduced a concerning degree of local dependency that would alter the conclusions we also ran all analysis on the baseline data which had the largest sample size of all the time points and to the 5 random samples. Evidence of local dependency introduced by repeated measures is often detected in the item residual correlation matrix when item-pairs have correlation values > 0.3 (Pallant and Tennant, 2007; Tennant and Conaghan, 2007; Marais, 2009; Andrich and Marais, 2019). Local dependency can also have an impact on reliability (either an increase or decrease) (Marais, 2009). Local dependency leading to an increased misfit to the Rasch model would reduce sample reliability and separation, whereas decreased misfit would increase sample reliability and separation (Marais, 2009). Finally, the Wilcoxon Rank Sum Test was used to compare the item locations of the full dataset and the baseline time point.

Correlation coefficients were calculated between the revised SAS total score (rSAS) derived from Rasch analysis for those constructs at the same ICF level, i.e., measures of depressive symptoms (polyserial correlation for ED-5D item anxiety/depression and Pearson correlation for SS-GDS). As these measures are theorized to relate to the same latent construct, the following criteria were used to qualify the strength: strong $\geq$ 0.8; moderate 0.50–0.80; weak < 0.50 (de Vet et al., 2011). Correlation coefficients were also calculated between gait speed (comfortable and maximum) and rSAS (apathy) and it is theorized not to be related.

For downstream outcomes (participation and health perception), linear regression was used with adjustment for age, sex and gait speed. Here the interpretation of the strength of correlation coefficients for novel relationships was used (strong: $\geq$ 0.5).

Linear regression was also used to estimate the extent to which the response of the stroke participants to the community-based participation-targeted intervention (from the original study) differed when measured using SAS and rSAS. The regression parameter for slope (β) quantified linear growth over time and the *t*-statistic, derived from the ratio of the β to its standard error (SE) was used for effect size. As regression models used explanatory variables with different measurement scales, to facilitate comparison standardized regression coefficients were used.

Generalized estimating equations (GEE) was used to make a direct comparison of change over time between the two measures. Here the value for apathy was the outcome and the explanatory variables were the version of the measure used to derive the score (original SAS and rSAS), time, and age and sex as adjustment variables. GEE considers the clustering of apathy value in the measure and adjusts the error variance accordingly.

RESULTS

Participant characteristics of the sample at baseline are summarized in **Table 1**. On average, participants had a stroke 2.4 years before study entry. **Table 2** summarizes the results of the Rasch analysis using data at all-time points. Disordered thresholds were observed in 8 of the 14 items. Disordered thresholds were consistently observed for "slightly" and "some" requiring the response options to be collapsed. Only item #8

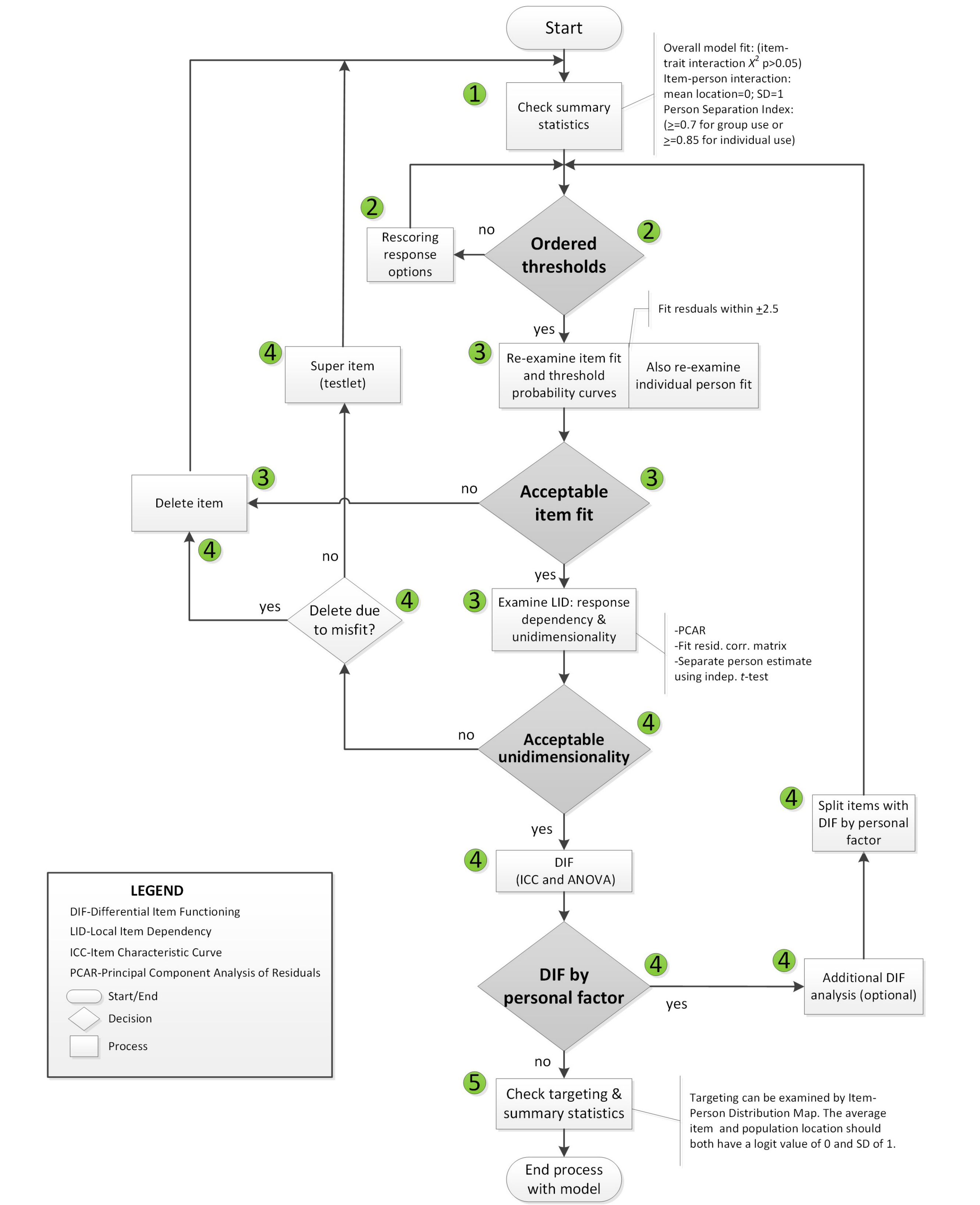

FIGURE 2 | Rasch analysis decision flowchart.

Step	Explanation
Step 1: Overall model fit (original data)	The items should line up hierarchically such that those items that need little ability to achieve the most optimal response level are at the low end and those items requiring more ability to achieve are higher. Overall goodness of model fit is indicated by a non-significant Chi-square test (p>0.05) after a Bonferroni adjustment for the number of items. Fit of each item and each person is as important, or even more important, than overall fit. Ideally, the person estimates from this measure should be centered on location 0 with a SD of 1 (targeting). Item and person fit is indicated when fit residual (deviance from pure linearity) values are within ±2.5 and the Chi-square test for fit is non-significant (>0.05). Those items that fail this criterion need to be looked at carefully to ensure their importance in scoring the latent trait. A fit residual of >+2.5 indicates the item does not fit the latent trait; a fit residual of <-2.5 indicates the item overfits and may be redundant. **Discrimination or person-separation:** This indicates how well people are differentiated by the spread of the item difficulty. The person-separation index (PSI) is interpreted like a Cronbach's alpha. The larger the index, the better is the discrimination which facilitates the measurement of change. Values of >0.9 are suitable for measuring within person change, values >0.7 are suitable for detecting group differences.
Step 2: Threshold order	There should be a logical ordering to the values that the person achieves such that achieving a more optimal response level should situate the person at a higher level of the latent trait. On these ordinal categories, a person with higher motivation (low apathy) is expected to achieve a higher response value. At lower motivation (high apathy) level, the person should achieve a lower response level. If the thresholds are disordered, the response categories need to be grouped to create ordered categories, sometimes reducing the responses to binary. The number of thresholds is equal to the number of response options—1 and reflects the number of "jumps" the person has to make for each item. This can be aided by visually inspecting the category probability curves (CPC) and the threshold probability curves (TPC).
Step 3: Item and person fit to model	Use criteria described in Step 1: A fit residual of >+2.5 indicates the item does not fit the latent trait; a fit residual of <-2.5 indicates the item overfits and may be redundant
Step 4: Unidimensionality and Local item dependency	**Unidimensionality:** A requirement of the Rasch model is that a single latent trait is being measured. This is assessed using a principal component analysis (PCA) of the fit residuals. The person-ability estimates derived from all pair-wise comparisons of the two most disparate set of items (those with the highest positive and negative loadings on the first factor) are compared using independent t-tests. For a set of items to be considered unidimensional, less than 5% of t values should be outside ± 1.96. When this value is greater than 5%, a binomial test of proportions is used to calculate the 95% confidence interval (CI) around the t-test estimate. Evidence of unidimensionality is still supported if the 5% value falls within the 95% CI. **Local item Dependency (LID):** Uniqueness of the information provided by the items is a requirement of the Rasch model. Items with pair-wise residual (after controlling for the latent trait) correlations greater than 0.3 could indicate lack of independence of the responses which inflates the reliability. Solutions include creating a super-item which combines the response options across items or choosing the one item that best suits the testing context.
Step 5: Differential item functioning (DIF)	The items should have the same ordering of difficulty across all people being measured defined by personal factors such as in this study, education or age. DIF is an indicator of item bias. Typically, DIF is indicated with a significant F-test from a two-way analysis of variance. A caution is that with large and sample sizes anything may be significant; with small sample sizes, nothing may be significant. Commonly used statistical packages provide a way of visually inspecting DIF (item characteristic curves are plotted by the level of each factor will support or not the information from the statistical approach). Two options are available for items with DIF, deletion or split scoring.
Step 6: Targeting and Overall model fit (final model)	An ideally targeted measure should include a set of items that spans the full range of the theoretical latent construct (-4 to +4 logits), and have a mean location of 0 with a standard deviation (SD) of 1. Ideally, the person estimates from this measure should be centered on location 0 with a SD of 1. Overall model fit (final model) – Same as Step 1

FIGURE 3 | Explanation of steps taken to fit the data to the Rasch model.

retained the original response options. The remaining items were rescored due to poor spacing between Threshold Probability Curves (TPC) or poor fit between the observed values and the expected TPC. Iterations of the Rasch analysis are summarized in **Table 3**. After rescoring, the 14 items SAS still did not fit the Rasch model. The overall chi-square value for item-trait interaction was statistically significant [$X^2(70) = 510.767$ p-value < 0.000] (**Table 3**). Item #3 and #5 had initial fit residuals of + 10.8 and + 5.9 respectively and were consistently > + 2.5 even after iterative adjustments with respect to the model and were deleted because of lack of fit to the hierarchical construct. Items #13 (fit residual of + 3.490) and #6 (fit residual of + 2.633) were iteratively deleted for the same reason. Item #11 (Are you unconcerned with many things?) did fit the Rasch model but had similar location (difficulty) values as item #10 (Are you indifferent to things?) with < 0.2 logit difference between the items and was deleted because it was poorly worded.

The remaining nine items met the requirements of the Rasch model with fit residuals between ± 2.5 as shown in **Table 2**. During iterative analysis, Item #7 and #14 had fit residuals < -2.5 indicating items over-discriminated the response pattern and may be redundant but in the final model the values were just within the cut-off (∼-2.4). In fact, item #14 fit residual improved to be within the set limit when item #11 was removed from the model.

Finally, The misfitting items (#3, #5, #6, #13) did not form a second factor based on the values of the PC Loadings when all 14 items were included. Additionally, the 4 items alone did not form a unidimensional construct (see **Supplementary Appendix**).

Local dependency was investigated using the correlation matrix of the residuals to examine response dependencies between items. All item-pair correlations were less than 0.3. A single construct, "apathy" was also supported with less than 5% of t-tests being significant (or the lower bound of the 95% binomial confidence interval should be less than 5%). Reliability based on the person separation index (PSI) is 0.74.

Targeting was assessed using the person-item threshold distribution map for the nine SAS items as shown in **Figure 4**. The figure shows item thresholds were reasonably well distributed over ∼5 logits with a near normal distribution (logit range ∼−2.0 to 2.8). The sample population had a logit mean of 1.60; SD:1.3 and a distribution range between ∼−1.6 and 4.4 logits showing some individuals at the low apathy end of the scale were not

TABLE 1 | Demographic and clinical characteristics of the sample (n = 238).

Characteristics	N	Mean (SD) or N (%) or median (IQR)
Age at stroke (years)	197	61.3 (12.3)
Age at interview (years)	214	63.1 (12.1)
Years since stroke	199	2.4 (2.2)
Women/Men	238	89/149 (37.4%/62.6%)
Inpatient rehabilitation	205	190 (92.7%)
Current smoker	218	23 (10.5%)
Has at least one co-morbidity	238	115 (48.3%)
Education (high school or less)	235	86 (36.6%)
Comfortable gait speed (m/s)	227	0.78 (0.41)
Max gait speed (m/s)	226	1.08 (0.59)
Depression SS-GDS (0–100, high is worst)	228	30.10 (26.62)
CHAMPS (hours)	236	25.82 (14.59)
RNL (0–100, high is better)	235	69.64 (21.78)

SD, Standard Deviation; [n]: # of participants that answered; IQR, interquartile range; SS-GDS, Stroke Specific Geriatric Depression Scale. Hours: Hours of meaningful activity (measured by CHAMPS) includes physical and social activities occurring outside of the home. RNL, Reintegration to Normal Living Index.

measured as reliably as there were no items that extended into that range. As shown in **Figure 4**, ∼13% of respondents' scores were higher than any item available for the assessment. Also, 30% of scores were > 2 SD above the standardized mean but only 2/9 items covered this range.

The summary statistics based on the final model with 9 items had an overall chi-square value for item-trait interaction that was still statistically significant [$X^2(45) = 89.103$ p-value < 0.0001] (**Table 3**).

As there are a large number of observations (n = 809) making the analysis overly sensitive to the detection of even trivial misfit adjusted sample sizes from the 5 random samples and the baseline time point were used. Each of the random samples and the baseline time point had non-significant overall chi-square p-values indicating global fit to the Rasch model. The average (the 5 random samples) p-value for fit of 0.28 and the average PSI of 0.73. The analysis of the baseline time point had a p-value for fit of 0.27 and PSI of 0.75 (see **Supplementary Appendix Table 1**).

Rasch analysis on each random sample and the baseline data point show that the fit residuals of the 9 items were within the set limit of ± 2.5. Item residual correlation matrix were all below the 0.3 cutoff indicating minimal local dependency in each of the random samples. PCA analysis within RUMM2020 did not indicate a second factor. The overall proportion of t-values falling outside a ± 1.96 range was less than 5% (or at least the lower bounds of 95% CIs of a binomial distribution were less than 5%).

There was no substantial DIF by time, age, sex, or education in the 5 random samples or the baseline data point. There was no discernable pattern of item DIF in the 5 random samples (see **Supplementary Appendix Table 1**). Graphically, there was no substantial deviation in the Item Characteristic Curves (ICC) for the personal factors (results not shown). Additionally, the results of the Wilcoxon Rank Sum Test showed that the full sample and the baseline time point did not differ on ranked item locations (p = 0.678).

The correlations between the original SAS or rSAS apathy total score with measures used to support interpretability are shown in **Table 4**. The strongest correlations (range −0.37 to −0.41) were observed between the convergent construct of depressed mood, measured using the EQ-5D anxiety/depression item and the SS-GDS. Neither apathy scale version was correlated with comfortable or maximum gait speed (m/s) with estimates ranging from 0.07 to 0.13.

Table 5 shows the extent to which the two different versions of the SAS (original and rSAS) explain downstream outcomes of participation and perceived health adjusted for age, sex, and gait speed. For CHAMPS-hours of meaningful activity, the effect of apathy was similar for the two versions (SAS β:0.32, t:4.85) and for (rSAS β:0.29, t:4.27). This similarity held for the other downstream outcomes. R^2 participation measures for SAS and rSAS ranged from 0.11 to 0.29, equivalent to correlations ranging

TABLE 2 | Starkstein's Apathy Scale items retained/deleted.

Item	Description	response options	Location	SE	Fit residual
I0001	Are you interested in learning new things?	0-1-1-2	−0.088	0.071	0.642
*I0002	Does anything interest you?	0-1-1-2	−0.065	0.070	−1.391
*I0003	3. Are you concerned about your condition?	0-1-1-2	Deleted-construct misfit		
I0004	Do you put much effort into things?	0-1-1-2	−0.179	0.072	0.037
*I0005	5. Are you always looking for something to do?	0-1-1-2	Deleted-construct misfit		
*I0006	6. Do you have plans and goals for the future?	0-1-1-2	Deleted-construct misfit		
I0007	7. Do you have motivation?	0-1-1-2	0.514	0.069	−2.482
I0008	Do you have the energy for daily activities?	0-1-2-3	0.740	0.053	0.805
*I0009	Does someone have to tell you what to do each day?	0-1-1-2	−0.596	0.073	2.017
I0010	Are you indifferent to things?	0-1-1-2	−0.180	0.070	0.305
I0011	11. Are you unconcerned with many things?	0-1-1-2	Location redundancy with #10, also poorly worded		
*I0012	Do you need a push to get started on things?	0-1-1-2	0.277	0.067	−0.174
*I0013	13. Are you neither happy nor sad, just in between?	0-1-1-2	Deleted-construct misfit		
*I0014	14. Would you consider yourself apathetic?	0-1-1-2	−0.423	0.070	−2.454

**Indicates items with disordered threshold requiring rescoring response options.*

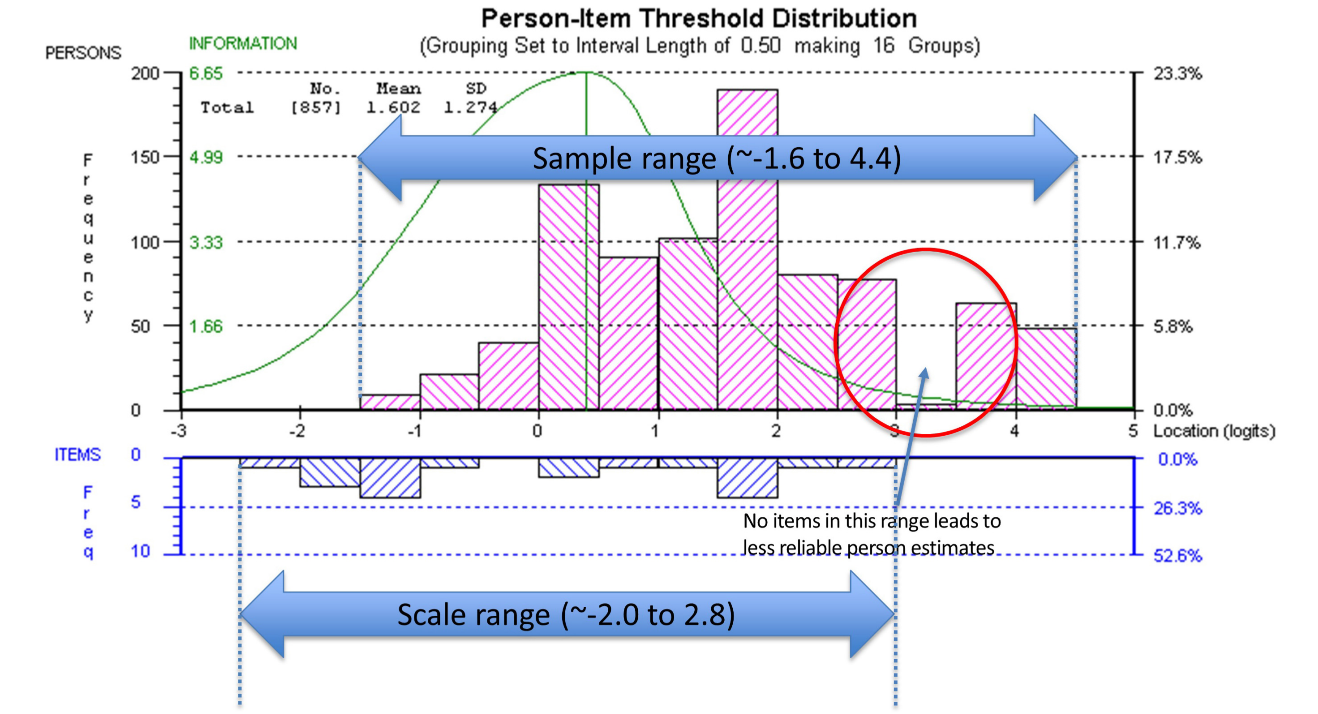

FIGURE 4 | Person-item threshold distribution map.

TABLE 3 | Summary fit statistics for Starkstein's Apathy Scale.

Model	Sample size (n)	Overall model fit	Item location	Item fit residual Mean (SD)	Person location	Person fit residual Mean (SD)	PSI	Unidimensionality *t*-tests %[a] (95%CI)
14 items	856	X^2 = 510.767, df = 70, *p* = 0.0000	0.0 (0.461)	−0.871 (3.635)	0.820 (0.695)	−0.166 (1.201)	0.74	–
14 items	856	X^2 = 422.249, df = 70, *p* = 0.0000	0.0 (0.678)	−0.003 (3.440)	1.124 (0.946)	−0.284 (1.263)	0.73	8.2 (6.5–10.2)
9 items	809	X^2 = 89.103, df = 45, *p* = 0.000099	0.0 (0.433)	−0.299 (1.523)	1.602 (1.274)	−0.333 (1.203)	0.74	4.9 (3.6–6.7)

Please note that the decrease in sample size in the 9 item model is due to the automatic removal of extreme scores. SD, standard deviation; PSI, person separation index; df, degrees of freedom. [a]*% of statistically significant t-tests.*

TABLE 4 | Relationship between rSAS apathy total score and measures used to support interpretability.

Correlation of the original SAS and rSAS score with known outcomes.

		Original SAS score				rSAS (9 items)		
Variable name	**N**	**Correlation**		***p*-value**		**Correlation**		***p*-value**
Anxiety/depression item EQ-5D	227	Spearman	−0.32	< 0.000	227	Polyserial	−0.40	< 0.000
SS-GDS (ladder 0–100)	217	Spearman	−0.39	< 0.000	228	Pearson	−0.40	< 0.000
Comfortable gait speed (m/s)	216	Spearman	0.08	0.270	227	Pearson	0.13	0.050
Max gait speed (m/s)	215	Spearman	0.05	0.497	226	Pearson	0.11	0.104

from 0.34 to 0.54 (considered moderate to strong); R^2 for health perception was approximately 0.25 (correlation 0.5, strong).

Table 6 shows key measurement properties of the two versions. GEE showed the 9-item rSAS (rescaled to be out of 42) produced a value that was 6.5 units lower than the full 14-item version with more variability. Floor and ceiling effects were minimal for the two versions. Estimated change over time was identical (0.12 units per month) which was not substantial (1.44 units per year) but greater than expected assuming no change (t: 3.6 and 2.4).

TABLE 5 | Regression of apathy (original total score or Rasch apathy score [rSAS]) as a predictor of downstream outcomes (participation/HRQL).

Participation (CHAMPS hours)		***N* = 208**					
***r* = 0.368; *R*² = 0.135 | F Change = 7.930 Sig. F Change = 0.000**				***r* = 0.338; *R*² = 0.114 F Change = 6.560 Sig. F Change = 0.000**			
	β	**t**	**Sig.**		**β**	**t**	**Sig.**
Original apathy score (SAS)	0.32	4.85	0.000	Rasch apathy score (rSAS)	0.29	4.27	0.000
Age at interview	−0.03	−0.03	0.643	Age at interview	−0.03	−0.44	0.658
Men(1) vs. Women(0)	0.07	1.12	0.266	Men(1) vs. Women(0)	0.07	1.05	0.293
Comfortable gait speed (m/s)	0.14	2.08	0.039	Comfortable gait speed (m/s)	0.13	1.93	0.054
Participation (RNL)		***N* = 207**					
***r* = 0.544; *R*² = 0.296 | F Change = 17.478 Sig. F Change = 0.000**				***r* = 0.526; *R*² = 0.277 | F Change = 19.351 Sig. F Change = 0.000**			
Model	**β**	**t**	**Sig.**	**Model**	**β**	**t**	**Sig.**
Original apathy score (SAS)	0.34	5.67	0.000	Rasch apathy score (rSAS)	0.31	5.11	0.000
Age at interview	0.24	4.03	0.000	Age at interview	0.24	3.93	0.000
Men(1) vs. Women(0)	−0.09	−1.56	0.121	Men(1) vs. Women(0)	−0.10	−1.61	0.109
Comfortable gait speed (m/s)	0.31	5.14	0.000	Comfortable gait speed (m/s)	0.30	4.89	0.000
Health perception (EQ-5D VAS)		***N* = 207**					
***r* = 0.492; *R*² = 0.242 | F Change = 16.086 Sig. F Change = 0.000**				***r* = 0.497; *R*² = 0.247 | F Change = 16.573 Sig. F Change = 0.000**			
	β	**t**	**Sig.**		**β**	**t**	**Sig.**
Original apathy score (SAS)	0.39	6.37	0.000	Rasch apathy score (rSAS)	0.40	6.51	0.000
Age at interview	0.14	2.21	0.028	Age at interview	0.13	2.13	0.034
Men(1) vs. Women(0)	0.02	0.36	0.719	Men(1) vs. Women(0)	0.02	0.33	0.742
Comfortable gait speed (m/s)	0.21	3.38	0.001	Comfortable gait speed (m/s)	0.19	3.08	0.002

TABLE 6 | Floor and Ceiling Effects and Responsiveness of the original SAS and rSAS.

***N* = 238**	**SAS**	**rSAS**
Items	14	9
Scoring range	0–42	0–19 (rescaled to 0–42)
Mean (SD) [baseline]	28.1 (6.5)	21.3 (9.2)
Adjusted difference*	−6.5 (SE:0.25; t:26.10)	
CV	0.23	0.43
Floor/Ceiling: n (%)	0/1 (0%/0.4%)	0/11 (0%/4.6%)
Change over time (β)	0.12	0.12
Effect size (β/SE)	3.6	2.4

**Adjusted difference (SE) estimated using GEE. CV, coefficient of variation; β, standardized beta; SD, standard deviation; SE, standard error.*

DISCUSSION

This study found that 9 of the original 14 items of the SAS fit a linear hierarchy (the Rasch model) suitable for measuring apathy in people with stroke. The results of the Rasch analysis on the original four-point ordinal scale showed that these thresholds were not used in a manner consistent with endorsing more positive response option with decreasing apathy (increasing motivation) (i.e., disordered thresholds). On 8 of the 14 items, participants consistently had difficulty differentiating reliably between the middle two response options "slightly" and "some." The rationale of having more response options is to try to be more precise in the assessment, but the observed disordered thresholds indicate that participants could not reliably distinguish between some choices, so this design choice was counter-productive. Not taking disordered thresholds into account can provide a false sense of reliability and increase measurement error.

The original SAS items measure several aspects of apathy according to Pedersen et al. (2012) such as: (1) diminished motivation (#7 and #12); (2) behavioral (#4, #5, #8, and #9); (3) cognitive (#1, #2, #6, #11); (4) emotional (#10 and #13); and (5) insight (#3 and #14). The study identified two factors using Exploratory Factor Analysis (EFA) that must be interpreted with caution since other authors have commented on unidimensionality and have concluded that EFA is not appropriate to test unidimensionali*ty* (Ziegler and Hagemann, 2015; Morita and Kannari, 2016).

Our analysis showed four items (#3, #5, #6, and #13) did not fit the latent construct and so were removed from the rSAS. The poor fit of item #3 "Are you concerned about your condition?" has also been described in three other Parkinson's studies as unreliable or ambiguous (Kirsch-Darrow et al., 2011; Pedersen et al., 2012; Morita and Kannari, 2016).

Morita and Kannari (2016) using Structured Equation Modeling Confirmatory Factor Analysis showed a single apathy construct but found that items #11 and #13 were not reflective of apathy in Parkinson's population (Morita and Kannari, 2016). Our analysis showed that item #13 "Are you neither happy nor sad, just in between?" did not fit the apathy construct and so was deleted from the measure. On the other hand, item #11

"Are you unconcerned with many things?" *did fit our model* but it was poorly worded. In using the SAS we had already noted patients having difficulty understanding the negative-positive phrasing "unconcerned/many" was difficult for people to interpret. Removing item #11 allowed item #14 "Would you consider yourself apathetic?" to remain anchoring the rSAS in the apathy domain. Additionally, item #10 had a similar location with 0.2 logit difference. Item #5 "Are you always looking for something to do?" and item #6 "Do you have plans and goals for the future?" did not fit the apathy construct in the stroke population, possibly because these are features of stroke rather than apathy.

Fit residuals for items #7 and #14 were just within the cut-off of −2.5, indicating marginal over-fit to the Rasch model. These were retained in the rSAS version. These two items are "reverse worded items" such that they might be considered essentially the same question with one positively worded and the other one negatively worded. Using words with opposite meaning has been discussed as a means to control for acquiescence and social desirability bias (van Sonderen et al., 2013) but also may have the effect of artificially increasing the reliability of the SAS.

Of the 9 items fitting the Rasch model, 7 are classified as self-report items as they query observable behaviors where the response provided by the patient could be amended based on other information independent of the patient. The only two PRO items in the rSAS version are item #7 and #14 asking the person directly if they think they are motivated or apathetic, respectively.

Our results did not show a correlation between apathy and tests of physical capacity (comfortable or maximum gait speed). A systematic review did report an association between apathy and increased disability post-stroke in most studies, however the disability outcomes were mostly related to activities of daily living which require effort. The authors also noted that they could not perform a quantitative meta-analysis due to the amount of heterogeneity in the outcomes and analyses used among the studies (van Dalen et al., 2013).

The literature supports depression as being distinct from apathy, but with some degree of overlap (Mayo et al., 2009; Clarke et al., 2011). Previous work also suggests that apathy can coexist with depression to varying degree in stroke populations specifically (Mayo et al., 2009). A review of several different apathy measures show low correlation with depression (Clarke et al., 2011). Our hypothesis was that depression constructs would correlate weakly with the rSAS. The weak correlation observed between rSAS and depression outcomes (SS-GDS and EQ-5D anxiety/depression) provided further evidence that apathy is a distinct construct from depression. The modest correlation (∼0.4) between SS-GDS and rSAS may still be due to the inclusion of overlapping items asking about loss of interest, doing new things, and energy level contained in both questionnaires. It would be prudent to select apathy and depression scales without items querying these common features to avoid misclassification, given that apathy and depression can coexist.

In our study, the amount of variance in participation measures explained by either the SAS or rSAS was small and very similar (range 0.114–0.296, **Table 5**). However, these 2 versions explained approximately 25% of the variance in health perception. This indicates that apathy has more to do with how people feel than what they do. What they do, may be influenced by family activities.

Our result is largely consistent with the reported estimates in the literature showing some association between apathy and HRQL in autoimmune, inherited, and neurodegenerative disorders; however, apathy was used as an outcome or as an exposure variable in the analysis in the different studies (Benito-León et al., 2012; Tang et al., 2014; Kamat et al., 2016; Fritz et al., 2018).

The distribution of the items of the SAS along the latent trait of apathy does not match the distribution of the values on the latent trait observed in sample. We have conceived of the apathy construct to range from apathy (low end of the scale) to motivation (high end of the scale) with the latent trait standardized to have a mean of 0 and a SD of 1. The mean location along the latent trait with all 14 items after rescoring due to disordered thresholds was 1.0 (SD 0.9); with the rSAS of 9-items the mean location was 1.6 (SD: 1.2). This means that people had more motivation than the items were able to measure suggesting that to adequately measure the full range of the construct other items would be needed. Until a stroke-specific measure of the apathy-motivation construct can be developed, researchers could use the nine rSAS items.

Sample Size Considerations

As a secondary analysis from an existing dataset we chose to use all the available data instead of selecting only one time point and test for DIF by time (of which there was no evidence). We used Bonferroni correction and/or *post hoc* downward random sample size adjustment when appropriate. Sample sizes between 250 and 500 are considered a good size for Rasch analysis of a well-targeted scale (Linacre, 1994; Chen et al., 2014; Hagell and Westergren, 2016); however, Type 1 error can occur with N as small as 200 (Müller, 2020).

When the additional time points are included in the dataset the question of introducing dependency arises. Wright argues that dependency probably occurs in a small way explaining that as "The patients are not identical patients. They have changed" (Wright, 2003). The patient answering the questionnaire will not have identical level of ability at different time points.

Additionally, the lack of independence in the observations owing to repeated measures does not affect item locations on the hierarchy. In fact, it is an advantage in that the effect of repeated measures (time) can be tested using differential item functioning (DIF), where the hypothesis is the ordering of the items is unaffected by time. This is a valuable psychometric property for the estimation of change and is used to distinguish between change from response shift and true change and also to identify if there is a learning effect (Guilleux et al., 2015).

Limitations

The data for this study came from an existing dataset generated from participants who were living with the long-term sequelae of stroke and were enrolled in a study of community based program to develop skilling to enhance community participation. As such, the measures available were fixed, and additional measures of apathy were not available. However, we have no evidence that the rSAS which is a subset of the items on the original SAS does not reflect the apathy construct as the correlation with measures of related constructs were closely similar for the two versions.

We feel the data generated here indicates that more work on the apathy construct is needed including qualitative work to generate items reflecting the content from the person's perspective, development of a robust scoring structure, and testing interpretability in diverse samples with respect to convergent and divergent constructs.

The rSAS was shown to be unidimensional according to the PCA of the residuals and the *t*-tests done on disparate items. Several approaches are suggested to test dimensionality of existing or new measures, including factor analysis and Rasch analysis which was used here. As factor analysis assumes the ordinal measurement scale is continuous and Rasch analysis converts ordinal scales to continuous, different conclusions about dimensionality can arise from these two approaches (Wright, 1996). Ideally, measures need to be constructed based on strong theoretical lines of an underlying unidimensional hierarchy (Waugh and Chapman, 2005). Future work on the apathy construct is warranted.

Usefulness of the Rasch Model

"Validity" is not a property of the measure, it is a property of the data with respect to how it can be interpreted. Rasch analysis tests the extent to which a set of items fit an underlying hierarchical model and so item and global fit support this hypothesis.Items that do not fit the Rasch model need to be investigated for sources of misfit which are often poor wording, a different construct, or negative vs. positive wording. Misfit indicates something is wrong with an item. Fit therefore indicates that the items collectively form a latent construct.

Future Directions

This study showed that a 9-item version of the original 14-item SAS could be used to assess apathy in chronic stroke patients without loss of content coverage but with gain in mathematical properties. The lack of targeting of the items to the range of motivation shown in this sample indicates additional items are needed. In addition, since the development of the original SAS, recommendations for developing PROs have been made (U.S. Department of Health and Human Services Food and Drug Administration, 2009). This process requires input from patients, caregivers, and clinicians and verification that all requirements for a robust, mathematically sound measure are met. Many disciplines are revisiting their legacy measures to assess the extent to which they measure up to these new standards and also adopting new measures that meet these new standards.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the McGill University Health Centre: Centre for Applied Ethics. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SH analyzed the data. SH and NM wrote the draft of the manuscript. All authors contributed to the interpretation of the results and writing the final version of the article.

FUNDING

SH was funded by the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.754103/full#supplementary-material

REFERENCES

Andrich, D., and Marais, I. (2019). *A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences*. Singapore: Springer Singapore, 482.

Benito-León, J., Cubo, E., Coronell, C., and Animo Study Group. (2012). Impact of apathy on health-related quality of life in recently diagnosed Parkinson's disease: The ANIMO study. *Mov. Disord.* 27, 211–218. doi: 10.1002/mds.23872

Carrozzino, D. (2019). Clinimetric approach to rating scales for the assessment of apathy in Parkinson's disease: A systematic review. *Prog. Neuro Psychopharmacol. Biol. Psychiatry* 94:109641. doi: 10.1016/j.pnpbp.2019.109641

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., and Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality Life Res.* 23, 485–493. doi: 10.1007/s11136-013-0487-5

Cinamon, J. S., Finch, L., Miller, S., Higgins, J., and Mayo, N. E. (2011). Preliminary evidence for the development of a stroke specific geriatric depression scale. *Int. J. Geriatr. Psychiatry* 26, 188–198. doi: 10.1002/gps.2513

Clarke, D. E., Ko, J. Y., Kuhl, E. A., van Reekum, R., Salvador, R., and Marin, R. S. (2011). Are the available apathy measures reliable and valid? A review of the psychometric evidence. *J. Psychosomat. Res.* 70, 73–97.

de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., and Knol, D. L. (2011). *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press.

EuroQol Group (2016). *EQ-5D*. Rotterdam: EuroQol Group.

Fritz, N. E., Boileau, N. R., Stout, J. C., Ready, R., Perlmutter, J. S., Paulsen, J. S., et al. (2018). Relationships Among Apathy, Health-Related Quality of Life, and Function in Huntington's Disease. *J. Neuropsychiatry Clin. Neurosci.* 30, 194–201.

Guilleux, A., Blanchin, M., Vanier, A., Guillemin, F., Falissard, B., Schwartz, C. E., et al. (2015). RespOnse Shift ALgorithm in Item response theory (ROSALI) for response shift detection with missing data in longitudinal patient-reported outcome studies. *Qual. Life Res.* 24, 553–564. doi: 10.1007/s11136-014-0876-4

Hagell, P., and Westergren, A. (2016). Sample Size and Statistical Conclusions from Tests of Fit to the Rasch Model According to the Rasch Unidimensional Measurement Model (Rumm) Program in Health Outcome Measurement. *J. Appl. Meas.* 17, 416–431.

Hansen, T., and Kjaersgaard, A. (2020). Item analysis of the Eating Assessment Tool (EAT-10) by the Rasch model: a secondary analysis of cross-sectional survey data obtained among community-dwelling elders. *Health Qual. Life Outcomes* 18:139. doi: 10.1186/s12955-020-01384-2

Hobart, J., and Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol. Assess.* 13, 1–177.

Kamat, R., Woods, S. P., Cameron, M. V., Iudicello, J. E., and Hiv Neurobehavioral Research Program (Hnrp) Group. (2016). Apathy is associated with lower mental and physical quality of life in persons infected with HIV. *Psychol. Health Med.* 21, 890–901. doi: 10.1080/13548506.2015.1131998

Kirsch-Darrow, L., Marsiske, M., Okun, M. S., Bauer, R., and Bowers, D. (2011). Apathy and Depression: Separate Factors in Parkinson's Disease. *J. Int. Neuropsychol. Soc.* 17, 1058–1066. doi: 10.1017/s1355617711001068

Le Heron, C., Holroyd, C. B., Salamone, J., and Husain, M. (2019). Brain mechanisms underlying apathy. *J. Neurol. Neurosurg. Psychiatry* 90, 302–312. doi: 10.1136/jnnp-2018-318265

Linacre, J. M. (1994). Sample Size and Item Calibration [or Person Measure] Stability. *Rasch Measure. Transact.* 7:328.

Linacre, J. M. (1999). Investigating rating scale category utility. *J. Outcome Meas.* 3, 103–122.

Marais, I. (2009). Response dependence and the measurement of change. *J. Appl. Meas.* 10, 17–29.

Marin, R. S. (1991). Apathy - a Neuropsychiatric Syndrome. *J. Neuropsych. Clin. N.* 3, 243–254. doi: 10.1176/jnp.3.3.243

Marin, R. S., Biedrzycki, R. C., and Firinciogullari, S. (1991). Reliability and validity of the apathy evaluation scale. *Psychiat. Res.* 38, 143–162. doi: 10.1016/0165-1781(91)90040-v

Mayo, N. E. (2015). *Dictionary of Quality of Life and Health Outcomes Measurement, Version 1*, 1 Edn. Milwaukee, WI: ISQOL.

Mayo, N. E., Anderson, S., Barclay, R., Cameron, J. I., Desrosiers, J., Eng, J. J., et al. (2015). Getting on with the rest of your life following stroke: a randomized trial of a complex intervention aimed at enhancing life participation post stroke. *Clin. Rehabil.* 29, 1198–1211. doi: 10.1177/0269215514565396

Mayo, N. E., Bronstein, D., Scott, S. C., Finch, L. E., and Miller, S. (2014). Necessary and sufficient causes of participation post-stroke: practical and philosophical perspectives. *Qual. Life Res.* 23, 39–47. doi: 10.1007/s11136-013-0441-6

Mayo, N. E., Fellows, L. K., Scott, S. C., Cameron, J., and Wood-Dauphinee, S. (2009). A Longitudinal View of Apathy and Its Impact After Stroke. *Stroke* 40, 3299–3307. doi: 10.1161/STROKEAHA.109.554410

Mayo, N. E., Figueiredo, S., Ahmed, S., and Bartlett, S. J. (2017). Montreal Accord on Patient-Reported Outcomes (PROs) use series – Paper 2: terminology proposed to measure what matters in health. *J. Clin. Epidemiol.* 89, 119–124. doi: 10.1016/j.jclinepi.2017.04.013

Morita, H., and Kannari, K. (2016). Reliability and validity assessment of an apathy scale for home-care patients with Parkinson's disease: a structural equation modeling analysis. *J. Phys. Ther. Sci.* 28, 1724–1727. doi: 10.1589/jpts.28.1724

Müller, M. (2020). Item fit statistics for Rasch analysis: can we trust them? *J. Statist. Distribut. Applicat.* 7:5.

Pallant, J. F., and Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *Br. J. Clin. Psychol.* 46, 1–18. doi: 10.1348/014466506x96931

Pallant, J., Miller, R., and Tennant, A. (2006). Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis. *BMC Psychiatry* 6:28. doi: 10.1186/1471-244X-6-28

Pedersen, K. F., Alves, G., Larsen, J. P., Tysnes, O.-B., Møller, S. G., and Brønnick, K. (2012). Psychometric Properties of the Starkstein Apathy Scale in Patients With Early Untreated Parkinson Disease. *Am. J. Geriatr. Psychiatry* 20, 142–148. doi: 10.1097/JGP.0b013e31823038f2

Petrillo, J., Cano, S. J., McLeod, L. D., and Coon, C. D. (2015). Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures: A Comparison of Worked Examples. *Value Health* 18, 25–34. doi: 10.1016/j.jval.2014.10.005

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institution for Educational Research.

Robert, P., Lanctot, K. L., Aguera-Ortiz, L., Aalten, P., Bremond, F., Defrancesco, M., et al. (2018). Is it time to revise the diagnostic criteria for apathy in brain disorders? The 2018 international consensus group. *Eur. Psychiat.* 54, 71–76. doi: 10.1016/j.eurpsy.2018.07.008

Starkstein, S. E., Fedoroff, J. P., Price, T. R., Leiguarda, R., and Robinson, R. G. (1993). Apathy following cerebrovascular lesions. *Stroke* 24, 1625–1630. doi: 10.1161/01.str.24.11.1625

Starkstein, S. E., Mayberg, H. S., Preziosi, T. J., Andrezejewski, P., Leiguarda, R., and Robinson, R. G. (1992). Reliability, validity, and clinical correlates of apathy in Parkinson's disease. *J. Neuropsych. Clin. Neurosci.* 4, 134–139. doi: 10.1176/jnp.4.2.134

Stewart, A. L., Mills, K. M., Kng, A. C., Haskell, W. L., Gillis, D., and Ritter, P. L. (2001). CHAMPS Physical Activity Questionnaire for Older Adults: outcomes for interventions. *Med. Sci. Sports Exerc.* 33, 1126–1141. doi: 10.1097/00005768-200107000-00010

Tang, W. K., Lau, C. G., Mok, V., Ungvari, G. S., and Wong, K. S. (2014). Apathy and health-related quality of life in stroke. *Arch. Phys. Med. Rehabil.* 95, 857–861.

Tennant, A., and Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care Res.* 57, 1358–1362. doi: 10.1002/art.23108

U.S. Department of Health and Human Services Food and Drug Administration (2009). *Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. Silver Spring, MD: U.S. Department of Health and Human Services Food and Drug Administration.

van Dalen, J. W., Moll van Charante, E. P., Nederkoorn, P. J., van Gool, W. A., and Richard, E. (2013). Poststroke apathy. *Stroke* 44, 851–860. doi: 10.1161/strokeaha.112.674614

van Sonderen, E., Sanderman, R., and Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: let's learn from cows in the rain. *PLoS One* 8:e68967–e. doi: 10.1371/journal.pone.0068967

Waugh, R. F., and Chapman, E. S. (2005). An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: what is the difference? Which method is better? *J. Appl. Meas.* 6, 80–99.

Wood-Dauphinee, S., and Williams, J. I. (1987). Reintegration to normal living as a proxy to quality of life. *J. Chronic Dis.* 40, 491–499. doi: 10.1016/0021-9681(87)90005-1

Wright, B. D. (1996). Comparing Rasch measurement and factor analysis. *Struct. Equat. Model. Multidiscipl. J.* 3, 3–24. doi: 10.1080/10705519609540026

Wright, B. D. (2003). Rack and Stack: Time 1 vs. Time 2. *Rasch Measure. Transact.* 17, 905–906.

Ziegler, M., and Hagemann, D. (2015). Testing the Unidimensionality of Items. *Eur. J. Psychol. Assess.* 31, 231–237. doi: 10.1027/1015-5759/a000309

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 04 January 2022
doi: 10.3389/fpsyg.2021.774703

School Burnout Inventory: Factorial Validity, Reliability, and Measurement Invariance in a Chilean Sample of High School Students

Marcos Carmona-Halty, Patricio Mena-Chamorro, Geraldy Sepúlveda-Páez and Rodrigo Ferrer-Urbina*

Escuela de Psicología y Filosofía, Universidad de Tarapacá, Arica, Chile

OPEN ACCESS

Edited by:
*Victor Zaia,
Faculdade de Medicina do ABC, Brazil*

Reviewed by:
*Katja Upadyaya,
University of Helsinki, Finland
İsmail Seçer,
Atatürk University, Turkey*

***Correspondence:**
*Marcos Carmona-Halty
mcarmonah@academicos.uta.cl*

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *12 September 2021*
Accepted: *22 November 2021*
Published: *04 January 2022*

Citation:
Carmona-Halty M, Mena-Chamorro P, Sepúlveda-Páez G and Ferrer-Urbina R (2022) School Burnout Inventory: Factorial Validity, Reliability, and Measurement Invariance in a Chilean Sample of High School Students. Front. Psychol. 12:774703. doi: 10.3389/fpsyg.2021.774703

This brief report assessed the psychometric validity and gender invariance of the School Burnout Inventory (SBI) –a measure of students' exhaustion, cynicism, and inadequacy– in a convenience sample of 972 high school Chilean students ranging between 12 and 18 years old. The results showed that: (1) the SBI produces adequate scores in terms of reliability; (2) two models (one solution of three related factors and one of second-order and three first-order factors) fitted adequately fit to our sample and was invariant across gender; and (3) the SBI scores were significantly related to other related constructs (i.e., study-related emotions, academic psychological capital, and academic engagement). Overall, the SBI was found to be a reliable and valid inventory to assess school burnout in Chilean high school students.

Keywords: school burnout, psychometric analyses, high school students, Chilean students, gender invariance

INTRODUCTION

The term burnout was first used in the late 70s to describe a state of exhaustion associated with human services professionals (Maslach and Jackson, 1981; Maslach et al., 2001; Schaufeli et al., 2002; Sakakibara et al., 2020). However, in the following years, this construct has become a social and scientific topic of great attention due to its extension to other domains, including students, and its impact on mental and physical health, achievement, and subsequent academic career (Fiorilli et al., 2014; Walburg, 2014; Walburg et al., 2016; Gabola et al., 2021). In this line, *school burnout* describes those students' who experience *emotional exhaustion* –characterized by feelings of strain and chronic fatigue resulting from overtaxing schoolwork, *cynicism* –characterized by an indifferent or a distal attitude toward schoolwork in general, a loss of interest in one's academic work, and not seeing it as meaningful, and *sense of inadequacy* –diminished feelings of competence and less achievement, and lack of accomplishment both in one's schoolwork and in school as a whole (Kiuru et al., 2008; Salmela-Aro et al., 2008, 2009b; Salmela-Aro and Upadyaya, 2014; Salmela-Aro, 2016; Salmela-Aro and Read, 2017).

Currently, we know that school burnout is directly related to school dropout (Bask and Salmela-Aro, 2013), boredom and neuroticism (Sulea et al., 2015), academic anxiety (Fiorilli et al., 2020), depressive symptoms (Moyano and Riaño-Hernández, 2013; Ponkosonsirilert et al., 2020), and suicidal ideas (Walburg et al., 2016). Conversely, it is inversely related to schoolwork engagement (Bresó et al., 2011; Salmela-Aro and Upadyaya, 2012), basic psychological needs (Sulea et al., 2015), well-being (Raiziene et al., 2013), and achievement (May et al., 2015; Gungor, 2019). Also, it has been found, based on the demands-resources model (Demerouti et al., 2001; Schaufeli and Bakker, 2004; Bakker and Demerouti, 2007), that study demands (e.g., schoolwork overload), and study resources (e.g., hope, efficacy, resilience, optimism) promote and prevent their appearance, respectively (Salanova et al., 2010; Salmela-Aro and Upadyaya, 2014; Gungor, 2019; Romano et al., 2021).

Research on the field has mostly been carried out using the School Burnout Inventory (SBI) developed by Salmela-Aro et al. (2009a). The SBI was initially validated in a Finnish context, demonstrating that school burnout can be explained both as a solution of three correlated factors (i.e., exhaustion, cynicism, and inadequacy) and by a second-order structure (i.e., school burnout) and three first-order factors (i.e., exhaustion, cynicism, and inadequacy). In the following years, the SBI has been validated in a few European and North American contexts (Spain–Moyano and Riaño-Hernández, 2013; Turkey–Secer et al., 2013; Italy–Fiorilli et al., 2014; France–Meylan et al., 2015; Switzerland–Gabola et al., 2021; Mexico–Osorio et al., 2019; and United States–May et al., 2020). Although the research mentioned above has led to significant advances in the study of school burnout, more research is warranted to assess the cross-cultural applicability of the SBI, especially in South American countries in which, to the best of our knowledge, no psychometric analyses have been conducted.

The current study aims to examine the psychometric properties of the SBI in a convenience sample of high school Chilean students. For this, we will follow a within-network and between-network construct validity. The first refers to assessing reliability, factor structure, and gender invariance, while the second refers to assessing the extent to which school burnout is associated with theoretically related constructs. In this line, we have selected three constructs that have been shown to play an important role in predicting school burnout. First, study-related emotions (e.g., Burr and Dallaghan, 2019; De la Fuente et al., 2020) are defined as those emotions that emerge in the educational environment and which relate to learning, achievement, and instructional processes that take place in the school (Pekrun et al., 2002). Second, academic psychological capital (e.g., Lupşa and Vîrgă, 2020; Vîrgă et al., 2020), defined as an individual's positive psychological state of development, characterized by hope, efficacy, resilience, and optimism (Luthans et al., 2015). Third, academic engagement (e.g., Salmela-Aro and Upadyaya, 2012; Sulea et al., 2015), defined as a positive, fulfilling, study-related state of mind characterized by vigor, dedication, and absorption (Schaufeli et al., 2002).

Based on the arguments presented, we hypothesize the following: The SBI will demonstrate acceptable psychometric properties in a Chilean sample of high school students. Also, we expect SBI scores to be direct and significantly related to study-related negative emotions, and inversely and significantly related to study-related positive emotions, academic psychological capital, and academic engagement. Together, the three constructs predict a significant percentage of the variance of school burnout.

MATERIALS AND METHODS

Participants

Nine hundred and seventy-two Chilean high school students (51% girls) in grades 5–12 (i.e., 12–18 years old, M = 14.41, SD = 1.635) participated in the study. They came from three different schools (each of them hosted approximately 550 students) located in two northern regions of the country (Arica y Parinacota and Tarapacá). Of 972 students, 16% were 12 years old, 17% were 13 years old, 17% were 14 years old, 20% were 15 years old, 19% were 16 years old, 8% were 17 years old, and 3% were 18 years old. In addition, 12% correspond to low, 81% to medium, and 7% to high socioeconomic levels.

Instrument

The SBI is a nine-item self-report scale grouped into three subscales: exhaustion at school (four items), cynicism toward the meaning of school (three items), and sense of inadequacy at school (two items). All items are scored on a six-point scale ranging from 1 (*completely disagree*) to 6 (*completely agree*).

Study–related emotions were measured using 12 items corresponding to two scales of positive (i.e., excited, energetic, inspired, at ease, relaxed, and satisfied) and negative (i.e., angry, anxious, disgusted, fatigued, discouraged, gloomy) emotions from the *Job–related Affective Well–being Scale* (Van Katwyk et al., 2000) adapted to the academic context (Carmona-Halty et al., 2019a). Students answered using a Likert–type with scores from 1 (*never*) to 5 (*always*), reflecting how they feel about their studies (for example, "my studies make me feel at ease" and "my studies make me feel anxious"). Cronbach's alpha and McDonald's omega for core study–related positive emotions were 0.736 and 0.734, respectively, and 0.713 and 0.718 for core study–related negative emotions.

Academic psychological capital was measured using the Academic Psychological Capital Questionnaire 12 (APCQ–12; Martínez et al., 2021). The APCQ–12 measure the four dimensions of the psychological capital construct –hope (four items; for example, "Right now I see myself as being pretty successful in my studies"), efficacy (three items; for example, "I feel sure when sharing information about my studies with other people"), resilience (three items; for example, "I usually take the stressful aspects of my studies in stride"), and optimism (two items; for example, "concerning my studies, I'm optimistic about what the future offers me")– using a Likert–type scale with scores from 1 (*totally disagree*) to 6 (*totally agree*). Cronbach's alpha and McDonald's omega for core academic psychological capital were 0.834 and 0.837, respectively.

Academic engagement was measured using the student version of the Utrecht Work Engagement Scale (UWES–9, Schaufeli et al., 2006) adapted to the Chilean context (Carmona-Halty et al., 2019b). The UWES measure the three dimensions of the academic engagement construct – vigor (three items; for example, “When I’m doing my work as a student, I feel bursting with energy”), dedication (three items; for example, “My studies inspire me”), and absorption (three items; for example, “I feel happy when I am studying intensely”)– using a Likert–type scale with scores from 0 (*never*) to 6 (*always*). Cronbach’s alpha and McDonald’s omega for core academic engagement were 0.852 and 0.855, respectively.

Procedure

The ethics committee of Universidad de Tarapacá approved the research project. The participants voluntarily filled out an online questionnaire during their usual class schedule. For the purposes of this study, a Spanish SBI version was developed following the International Test Commission Guidelines for test translation and adaptation (Muñiz et al., 2013). Prior to general data collection, the SBI items were piloted on a small sample ($n = 12$) of high school students to find difficulties in understanding the translated and adapted version of the inventory. At this stage, no students reported difficulties in understanding the items. Sampling was non-probabilistic by convenience (Ato et al., 2013) and no participant was removed from dataset.

Statistical Analyses

All data analyses were conducted using Jamovi 1.2 (The Jamovi Project, 2020) and Mplus 8.2 (Muthén and Muthén, 1998). For preliminary analysis, descriptive statistics, normality tests, and gender differences were tested. For reliability analysis, Cronbach’s alpha and McDonald’s omega coefficients, corrected homogeneity index, and test-retest (with a time lag of four-month) index were calculated. For evidence of validity based on the internal structure, a confirmatory factor analysis (CFA) was realized, using the Robust Maximum Likelihood (RML) estimation method. In addition, gender invariance through multi-group CFA and three levels of equivalence were assessed (i.e., configural invariance, metric invariance, and scalar invariance) using changes in CFI and Standard Root Mean Residual (SRMR; $\Delta < 0.010$) as criteria for determining whether measurement invariance was established (Cheung and Rensvold, 2002; Chen, 2007; Dimitrov, 2010). Schreiber (2017) recommendations were used to help evaluate the cut-off and determine model fit for the following indicators: chi-squared (χ^2) and normed χ^2, root-mean-squared error of approximation (RMSEA) with a confidence interval (90% CI), comparative fit index (CFI), Tucker-Lewis index (TLI), and SRMR. Finally, to examine criterion validity of SBI, we conducted both correlation and regression analysis between school burnout, study–related (positive and negative) emotions,

TABLE 1 | Descriptive information of the school burnout inventory.

	Descriptive statistics				
	Mean (SD)	**S**	**K**	**Shapiro-Wilk (*p*)**	**Corrected homogeneity index**
1. I feel overwhelmed by my schoolwork.	3.84 (1.37)	−0.16	−0.77	0.93 (> 0.001)	0.39
2. I feel a lack of motivation in my schoolwork and often think of giving up.	3.24 (1.67)	0.12	−1.21	0.90 (> 0.001)	0.56
3. I often have feelings of inadequacy in my schoolwork.	3.11 (1.35)	0.16	−0.58	0.93 (> 0.001)	0.56
4. I often sleep badly because of matters related to my schoolwork.	3.38 (1.74)	0.04	−1.31	0.89 (> 0.001)	0.51
5. I feel that I am losing interest in my schoolwork.	2.89 (1.58)	0.46	−0.92	0.89 (> 0.001)	0.58
6. I’m continually wondering whether my schoolwork has any meaning.	3.57 (1.73)	−0.08	−1.30	0.90 (> 0.001)	0.43
7. I brood over matters related to my schoolwork a lot during my free time.	3.29 (1.42)	0.25	−0.67	0.92 (> 0.001)	0.10
8. I used to have higher expectations of my schoolwork than I do now.	3.66 (1.65)	−0.12	−1.14	0.91(> 0.001)	0.48
9. The pressure of my schoolwork causes me problems in my close relationships with others.	3.02 (1.71)	0.31	−1.20	0.88 (> 0.001)	0.54

SD = Standard Deviation; S = Skewness; K = Kurtosis.

TABLE 2 | Fit indexes for single–group and multiple–group CFA of the school burnout inventory.

	χ^2	*df*	χ^2/df	RMSEA	90% CI	CFI	TLI	SRMR	CMs	Δ CFI	Δ SRMR
Single–group CFA											
M1 One factor	181.42	20	9.071	0.091	[0.079,0.104]	0.898	0.857	0.049	–	–	–
M2 Three–factors	75.456	17	4.438	0.059	[0.046,0.073]	0.963	0.939	0.030	–	–	–
M3 Second order factor	75.457	17	4.438	0.059	[0.046,0.073]	0.963	0.939	0.030	–	–	–
Multiple–group CFA											
M4 Configural invariance	141.64	34	4.165	0.081	[0.067,0.095]	0.975	0.958	0.027	–	–	–
M5 Metric invariance	141.55	39	3.629	0.074	[0.061,0.087]	0.976	0.965	0.028	M4–M5	0.001	0.001
M6 Scalar invariance	159.50	68	2.345	0.053	[0.042,0.063]	0.978	0.982	0.030	M5–M6	0.003	0.003

χ2 = Chi-square; df = degree of freedom; RMSEA = Root Mean Square Error of Approximation; CI = 90% Confidence Interval; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMR = Standardized Root Mean Square Residual; CMs = Comparisons between models.

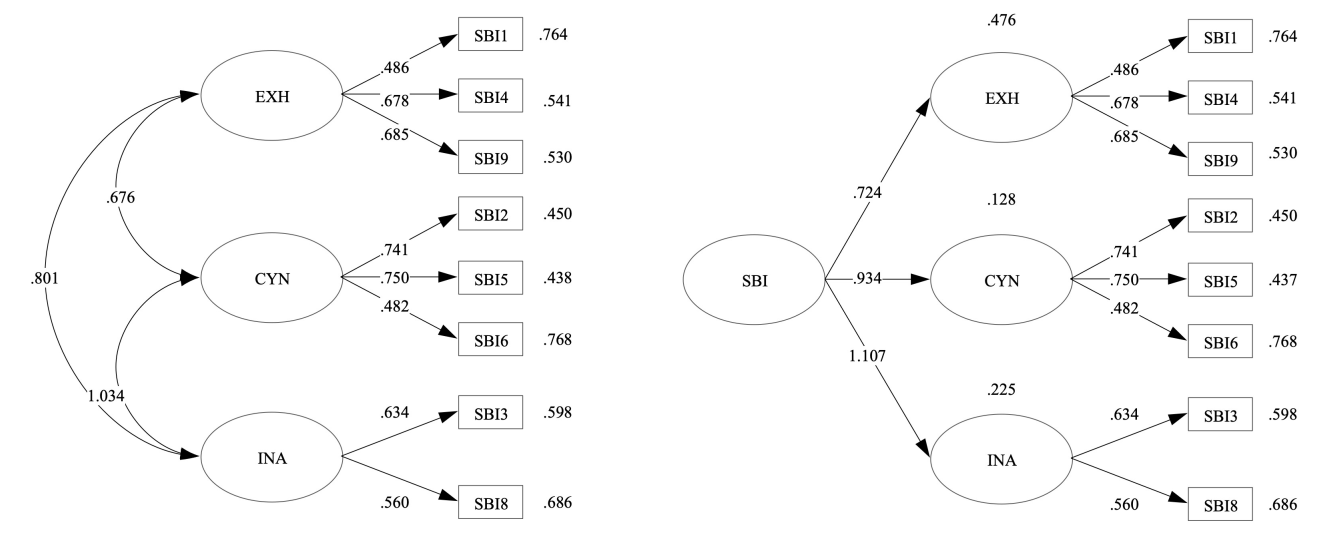

FIGURE 1 | Measurement models of school burnout.

TABLE 3 | Correlations and regression analysis.

	Pearson's correlation indices				Linear regression indices				
	1	**2**	**3**	**4**	**R^2**	**F**	**β**	**β_{std}**	***t***
1. School burnout	–				0.291	99.206			
2. Study–related positive emotions	−0.284*	–					0.037	0.026*ns*	0.624
3. Study–related negative emotions	0.485*	−0.275*	–				0.633	0.396*	13.661
4. Academic psychological capital	−0.349*	0.619*	−0.296*	–			−0.157	−0.146*	−3.852
5. Academic engagement	−0.368*	0.742*	−0.352*	0.677*			−0.123	−0.149*	−3.304

**p < 0.001; ns = non-significant.*

academic psychological capital (i.e., hope, efficacy, resilience, and optimism), and academic engagement (i.e., vigor, dedication, and absorption).

RESULTS

Table 1 shows the descriptive statistics for the SBI at item levels. Independent sample t-test revealed that there are not statistical significance differences between boys' (M = 3.309, SD = 0.949) and girls' (M = 3.354, SD = 0.960) SBI scores [t (970) = 0.728, $p > 0.05$, d = 0.047, 95% IC (-0.165,0.076)].

Within-Network Construct Validation

Considering the corrected homogeneity index, it seems convenient to eliminate item seven (see **Table 1**). Thus, this item was excluded from further analyses. The SBI showed good internal consistency with Cronbach's alpha and McDonald's omega indexes of 0.776 and 0.787, respectively. In addition, internal consistency for each dimension were at least sufficient in both exhaustion (ω = 0.65; α = 0.64) and cynicism (ω = 0.70; α = 0.67) dimensions, although slightly lower in the case of inadequacy dimension (ω = 0.52; α = 0.51).

According to previous research on the evidence of the validity of the SBI (e.g., Salmela-Aro et al., 2009a), three models were examined, namely, one that assumes that there is one latent factor underlying all the SBI items (M1), one that proposes three related factors (M2), and one that explains three first-order factors by a second-order factor (M3). The results of the CFA show that M2 and M3 showed an adequate fit (see **Table 2** and **Figure 1**). The multiple-group CFA shows that the differences in the CFI and SRMR across the three invariance models (i.e., configural, metric, and scalar) were lower than .01, which indicates gender invariance (see **Table 2**). Finally, in terms of test-retest reliability (n = 775), the total SBI score shows a correlation of 0.597, while its dimensions reach similar values: exhaustion (r = 0.541; $p < 0.001$), cynicism (r = 0.548; $p < 0.001$), inadequacy (r = 0.501; $p < 0.001$).

Between-Network Construct Validation

As shown in **Table 3**, our results showed that school burnout was significant –and in the expected direction– related with all assessed constructs. In addition, they predict a substantial percentage (29.10%) of the school burnout variance measured with the SBI.

DISCUSSION

The current study aims to examine the psychometric properties of the SBI in a convenience sample of high school Chilean students to offer a potential contribution to the school burnout literature.

First, our results suggest that both the three–factors solution and second–order factor of SBI was applicable for high school Chilean students and comparable across gender, which is coherent with previous research (Salmela-Aro et al., 2009a; Moyano and Riaño-Hernández, 2013; Secer et al., 2013; Fiorilli et al., 2014; Meylan et al., 2015; Osorio et al., 2019; May et al., 2020). In addition, reliability analysis, based on Cronbach's alpha and McDonald's omega indexes, indicated good internal consistency, whereas the test–retest index shows a moderated relation over a time of four months.

Second, our results suggest that the SBI score relates to other theoretically linked variables to the school burnout construct. In this line, students with high levels of exhaustion, cynicism, and inadequacy, on the one hand, are more likely to experience a high frequency of student–related negative emotions (i.e., angry, anxious, disgusted, fatigued, discouraged, and gloomy), and, on the other hand, a low frequency of study–related positive emotions (i.e., excited, energetic, inspired, at ease, relaxed, and satisfied). In addition, they will also report lower levels of academic, personal resources –measured in terms of hope, efficacy, resilience, and optimism (i.e., psychological capital)– and academic well–being –measured in terms of vigor, dedication, and absorption (i.e., academic engagement)– which is coherent with previous studies (Salmela-Aro and Upadyaya, 2012; Raiziene et al., 2013; Burr and Dallaghan, 2019; Gungor, 2019; Fiorilli et al., 2020; Romano et al., 2021).

Third, our study has some important strengths, such as the large sample used, considering both within-network and between-network approaches, and analyzing the temporal stability of the SBI. However, as a limitation, we can mention that we did not obtain representative data from high school Chilean students; thus, the generalization of the results should be made with caution. Additionally, we use only self–report measures without considering other sources of information (e.g., teachers 'reports or grade point average).

Finally, future research can expand our results to other student segments (e.g., undergraduate university students), use samples from different South American countries, and propose cut-off criteria for low, medium, and high levels of school burnout.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comité Ético Científico of the Universidad de Tarapacá. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

All authors contributed equally to the research design and wrote the manuscript.

FUNDING

This work was supported by Universidad de Tarapacá, Chile (UTA Mayor Award Number 3764–20).

REFERENCES

Ato, M., López-García, J., and Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología* 29, 1038–1059. doi: 10.6018/analesps.29.3.178511

Bakker, A. B., and Demerouti, E. (2007). The job demands-resources model: State of the art. *J. Manag. Psychol.* 22, 309–328. doi: 10.1108/02683940710733115

Bask, M., and Salmela-Aro, K. (2013). Burned out to drop out: Exploring the relationship between school burnout and school dropout. *Eur. J. Psychol. Edu.* 28, 511–528. doi: 10.1007/s10212-012-0126-5

Bresó, E., Schaufeli, W. B., and Salanova, M. (2011). Can a self-efficacy-based intervention decrease burnout, increase engagement, and enhance performance? A quasi-experimental study. *High. Edu.* 61, 339–355. doi: 10.1007/s10734-010-9334-6

Burr, J., and Dallaghan, G. L. B. (2019). The relationship of emotions and burnout to medical students' academic performance. *Teach. Learn. Med.* 2019:237. doi: 10.1080/10401334.2019.1613237

Carmona-Halty, M., Salanova, M., Llorens, S., and Schaufeli, W. B. (2019a). How psychological capital mediates between study-related positive emotions and academic performance. *J. Hap. Stud.* 20, 605–617. doi: 10.1007/s10902-018-9963-5

Carmona-Halty, M., Schaufeli, W. B., and Salanova, M. (2019b). The Utrecht Work Engagement Scale for Students (UWES–9S): psychometric properties and measurement invariance in a Chilean sample of undergraduate university students. *Front. Psychol.* 10:1017. doi: 10.3389/fpsyg.2019.01017

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Model. Multidis. J.* 14, 464–504. doi: 10.1080/10705510701301834

Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Model.* 9, 233–255. doi: 10.1207/s15328007sem0902_5

De la Fuente, J., Paoloni, P. V., Vera-Martinez, M. M., and GarzonUmerenkova, A. (2020). Effects of levels of self-regulation and situational stress on achievement emotions in undergraduate students: class, study, and testing. *Int. J. Environ. Res. Public Health* 17:4293. doi: 10.3390/ijerph17124293

Demerouti, E., Bakker, A. B., Nachreiner, F., and Schaufeli, W. B. (2001). The job demands-resources model of burnout. *J. Appl. Psychol.* 86, 499–512. doi: 10.1037/0021-9010.86.3.499

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measur. Eval. Counsel. Dev.* 43, 121–149. doi: 10.1177/0748175610373459

Fiorilli, C., Farina, E., Buonomo, I., Costa, S., Romano, L., Larcan, R., et al. (2020). Trait emotional intelligence and school burnout: The mediating role of resilience and academic anxiety in high school. *Int. J. Environ. Res. Public Health* 17:3058. doi: 10.3390/ijerph17093058

Fiorilli, C., Galimberti, V., De Stasio, S., Di Chiacchio, C., and Albanese, O. (2014). L'utilizzazione dello School Burnout Inventory (SBI) con student italiani di scuola superior di primo e secondo grado. *Psicologia Clinica Dello Sviluppo* 18, 403–423. doi: 10.1449/78365

Gabola, P., Meylan, N., Hoscoët, M., De Stasio, S., and Fiorilli, C. (2021). Adolescents' school burnout: a comparative study between Italy and Switzerland. *Eur. J. Invest. Health Psychol. Edu.* 11, 849–859. doi: 10.3390/ejihpe11030062

Gungor, A. (2019). Investigating the relationship between social support and school burnout in Turkish middle school students: the mediating role of hope. *Sch. Psychol. Int.* 40, 581–597. doi: 10.1177/0143034319866492

Kiuru, N., Aunola, K., Nurmi, J. E., Leskinen, E., and Salmela-Aro, K. (2008). Peer group influence and selection in adolescents' school burnout: A longitudinal study. *Merril-Palmer Quart.* 54, 23–55. doi: 10.1353/mpq.2008.0008

Lupşa, D., and Vîrgă, D. (2020). Psychological capital, health, and performance: The mediating role of burnout. *Psihologia Resurselor Umane* 18, 7–22. doi: 10.24837/pru.v18i1.458

Luthans, F., Youssef-Morgan, C. M., and Avolio, B. (2015). *Psychological Capital and Beyond*. New York, NY: Oxford University Press.

Martínez, I. M., Meneghel, I., Carmona-Halty, M., and Yousef-Morgan, C. (2021). Adaptation and validation to spanish of the psychological capital questionnaire – 12 (PCQ–12) in academic contexts. *Curr. Psychol.* 40, 3409–3416. doi: 10.1007/s12144-019-00276-z

Maslach, C., and Jackson, S. E. (1981). The measurement of experienced burnout. *J. Org. Behav.* 2, 99–133. doi: 10.1002/job.4030020205

Maslach, C., Schaufeli, W. B., and Leiter, M. P. (2001). Job burnout. *Ann. Rev. Psychol.* 52, 397–422. doi: 10.1146/annurev.psych.52.1.397

May, R. W., Bauer, K. N., and Fincham, F. D. (2015). School burnout: Diminished academic and cognitive performance. *Learn. Ind. Diff.* 42, 126–131. doi: 10.1016/j.lindif.2015.07.015

May, R. W., Bauer, K. N., Seibert, G. S., Jaurequi, M. E., and Fincham, F. D. (2020). School burnout is related to sleep quality and perseverative cognition regulation at bedtime in young adults. *Learn. Ind. Diff.* 78:101821. doi: 10.1016/j.lindif.2020.101821

Meylan, N., Doudin, P. A., Antonietti, J. P., and Stephan, P. (2015). School Burnout Inventory: a French validation. *Eur. Rev. Appl. Psychol.* 65, 285–293. doi: 10.1016/j.erap.2015.10.002

Moyano, N., and Riaño-Hernández, D. (2013). Burnout in spanish adolescents: adaptation and validation of the school burnout inventory. *Ansiedad Estrés* 19, 95–103. doi: 10.1186/1472-6920-11-103

Muñiz, J., Elosua, P., and Hambleton, R. K. (2013). International test commission guidelines for test translation and adaption: Second edition. *Psicothema* 25, 151–157. doi: 10.7334/psicothema2013.24

Muthén, L., and Muthén, B. (1998). *Mplus User's Guide (8th Edition)*. Los Angeles: Muthén & Muthén.

Osorio, M., Prado, C., Parrello, S., and Bazán, G. (2019). Psychometric characteristics and factor structure of the School Burnout Inventory in Mexican university Students. *Revista Iberoamericana de Diagnóstico y Evaluación* 2, 141–150. doi: 10.21865/RIDEP55.2.10

Pekrun, R., Goetz, T., and Titz, W. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Edu. Psychol.* 37, 91–115. doi: 10.1207/S15326985EP3702_4

Ponkosonsirilert, T., Laemsak, O., Pisitsungkagarn, K., Jarukasemthawee, S., Audboon, S., and Leangsuksant, T. (2020). Stress, self-compassion, and school burnout in Thai high school students. *Int. J. Adol. Med. Health* 2020:109. doi: 10.1515/ijamh-2020-0109

Raiziene, S., Pilkauskaite-Valickiene, R., and Zukauskiene, R. (2013). School burnout and subjective well-being: Evidence from cross-lagged relations in a 1 year longitudinal sample. *Proc. Soc. Behav. Sci.* 116, 3254–3258. doi: 10.1016/j.sbspro.2014.01.743

Romano, L., Consiglio, P., Angelini, G., and Fiorilli, C. (2021). Between academic resilience and burnout: The moderating role of satisfaction on school context relationships. *Eur. J. Invest. Health Psychol. Edu.* 11, 770–780. doi: 10.3390/ejihpe11030055

Sakakibara, K., Shimazu, A., Toyama, H., and Schaufeli, W. B. (2020). Validation of the japanese version of the burnout assessment tool. *Front. Psychol.* 11:1819. doi: 10.3389/fpsyg.2020.01819

Salanova, M., Schaufeli, W. B., Martinez, I., and Bresó, E. (2010). How obstacles and facilitators predict academic performance: the mediating role of study burnout and engagement. *Anx. Stress Cop.* 23, 53–70. doi: 10.1080/10615800802609965

Salmela-Aro, K. (2016). Dark and bright sides of thriving – school burnout and engagement in the Finnish context. *Eur. J. Dev. Psychol.* 2016:517. doi: 10.1080/17405629.2016.1207517

Salmela-Aro, K., and Read, S. (2017). Study engagement and burnout profiles among Finnish higher education students. *Burnout Res.* 7, 21–28. doi: 10.1016/j.burn.2017.11.001

Salmela-Aro, K., and Upadyaya, K. (2012). The schoolwork engagement inventory. *Eur. J. Psychol. Assess.* 28, 60–67. doi: 10.1027/1015-5759/a000091

Salmela-Aro, K., and Upadyaya, K. (2014). School burnout and engagement in the context of demands-resources model. *Br. J. Edu. Psychol.* 84, 137–151. doi: 10.1027/1016-9040.13.1.12

Salmela-Aro, K., Kiuru, N., Pietikäinen, M., and Jokela, J. (2008). Does school matter? The role of school context in adolescents' school-related burnout. *Eur. Psychol.* 13, 12–23.

Salmela-Aro, K., Savolainen, H., and Holopainen, L. (2009b). Depressive symptoms and school burnout during adolescence: evidence from two cross-lagged longitudinal studies. *J. Youth Adol.* 38, 1316–1327. doi: 10.1007/s10964-008-9334-3

Salmela-Aro, K., Kiuru, N., Leskinen, E., and Nurmi, J. E. (2009a). School burnout inventory (SBI): reliability and validity. *Eur. J. Psychol. Assess.* 25, 48–57. doi: 10.1027/1015-5759.25.1.48

Schaufeli, W. B., and Bakker, A. B. (2004). Job demands, job resources and their relationship with burnout and engagement: a multi-sample study. *J. Org. Behav.* 25, 293–299. doi: 10.1002/job.248

Schaufeli, W. B., Bakker, A. B., and Salanova, M. (2006). The measures of work engagement with a short questionnaire: A cross-national study. *Edu. Psychol. Measur.* 66, 701–716. doi: 10.1177/0013164405282471

Schaufeli, W. B., Martínez, I., Marques Pinto, A., Salanova, M., and Bakker, A. B. (2002). Burnout and engagement in university students: A cross-national study. *J. Cross-Cultural Psychol.* 33, 464–481. doi: 10.1177/0022022102033005003

Schreiber, J. B. (2017). Update to core reporting practices in structural equation modeling. *Res. Soc. Adm. Pharm.* 13, 634–643. doi: 10.1016/j.sapharm.2016.06.006

Secer, I., Halmatov, S., Veyis, F., and Ates, B. (2013). Adapting school burnout inventory to turkish culture: study of validity and reliability. *Turk. J. Edu.* 2, 13–24.

Sulea, C., van Beek, I., Sarbescu, P., Virga, D., and Schaufeli, W. B. (2015). Engagement, boredom, and burnout among students: Basic need satisfaction matters more than personality traits. *Learn. Ind. Diff.* 42, 132–138.

The Jamovi Project (2020). *Jamovi (version 1.8.1) [Computer Software]*. Available online at: https://www.jamovi.org. [accessed 5 April 2021]

Van Katwyk, P. T., Fox, S., Spector, P. E., and Kelloway, E. K. (2000). Using the Job–related Affective Well–being scale (JAWS) to investigate affective responses to work stressors. *J. Occup. Health Psychol.* 5, 219–230. doi: 10.1037/1076-8998.5.2.219

Vîrgă, D., Pattusamy, M., and Kumar, D. P. (2020). How psychological capital is related to academic performance, burnout, and boredom? The mediating role of study engagement. *Curr. Psychol.* 2020:9. doi: 10.1007/s12144-020-01162-9

Walburg, V. (2014). Burnout among high school students: A literature review. *Child. Youth Serv. Rev.* 42, 28–33. doi: 10.1016/j.chilyouth.2014.03.020

Walburg, V., Mialhes, A., and Moncla, D. (2016). Does school-related burnout influence problematic Facebook use? *Child. Youth Serv. Rev.* 61, 327–331. doi: 10.1016/j.childyouth.2016.01.009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 January 2022
doi: 10.3389/fpsyg.2021.774847

Social Networks Addiction (SNA-6) – Short: Validity of Measurement in Mexican Youths

*Edwin Salas-Blas[1], César Merino-Soto[1], Berenice Pérez-Amezcua[2] and Filiberto Toledano-Toledano[3,4]**

[1] Instituto de Investigación de Psicología, Universidad de San Martín de Porres, Lima, Peru, [2] Centro de Investigación Transdisciplinar en Psicología, Universidad Autónoma del Estado de Morelos, Cuernavaca, Mexico, [3] Unida de Investigación en Medicina Basada en Evidencias, Hospital Infantil de México Federico Gómez, Mexico City, Mexico, [4] Unidad de Investigación Sociomédica, Instituto Nacional de Rehabilitación Luis Guillermo Ibarra Ibarra, Mexico City, Mexico

OPEN ACCESS

Edited by:
Paola Gremigni,
University of Bologna, Italy

Reviewed by:
André Luiz Monezi Andrade,
Pontifical Catholic University of Campinas, Brazil
Ana María Ruiz-Ruano García,
University of Granada, Spain

Correspondence:
Filiberto Toledano-Toledano
filiberto.toledano.phd@gmail.com

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *13 September 2021*
Accepted: *15 December 2021*
Published: *12 January 2022*

Citation:
Salas-Blas E, Merino-Soto C, Pérez-Amezcua B and Toledano-Toledano F (2022) Social Networks Addiction (SNA-6) – Short: Validity of Measurement in Mexican Youths. Front. Psychol. 12:774847. doi: 10.3389/fpsyg.2021.774847

The excessive use of social networks needs to be addressed, and this phenomenon needs to be measured for the purpose of evaluation, prevention, and intervention among adolescents and young people. The objective of the study was to adapt and psychometrically validate the Brief Scale of Addiction to Social Networks (SNA-6) among Mexican adolescents and young adults. The participating sample consisted of 2,789 students from 6 public educational campuses in Cuernavaca (Morelos, Mexico). Data collection was carried out through a web platform to strictly maintain anonymity, voluntary participation, and confidentiality. Data analysis first focused on the detection of possible response biases (random intercept model and careless/insufficient effort), the quality of the response structure partial credit model (PCM), dimensionality (CFA and invariance), and the relationship with external variables. It was found that when the range of efficient response options was limited to less than five, reliability was high (0.91), and unidimensionality was maintained. Response biases slightly affected the dimensional structure of the instrument. Measurement invariance reached scalar invariance in the sex, age, and campus groups. The association with sensation seeking and depression, controlling for sex and age covariates, was statistically significant, small, and theoretically consistent. Implications of the results are discussed.

Keywords: random intercept model, social networks, addiction, validity, adolescents, Mexican

INTRODUCTION

A Digital 2021 report carried out by We Are Social and Hootsuite indicates that the number of users of social networks (SNs) in the world continues to increase vigorously,[1] maintaining that by January 2021, there were 4,200 million users, which is approximately 53% of the world population. This represents 13% growth over the previous year. In the Americas, the countries that have experienced the greatest growth are Mexico, the United States, and Brazil. The largest group of internet users are between 25 and 34 years old, followed by those between 18 and 24 years old. In both age groups,

[1] https://datareportal.com/reports/digital-2021-global-overview-report

men spend more time networking. The average time invested in SNs globally is 2 h 25 min, but in Mexico, it is 3 h.

The development of information technologies, particularly smartphones, as well as the widespread use of the internet and its current low costs, are elements that could explain the increase in the use of virtual SNs, which are becoming increasingly attractive to consumers. However, SNs have positive aspects, such as allowing people to establish rapid communication over distances and being used in work or study situations. However, if SNs are not used properly, they could result in health problems (Salas, 2014; Loss et al., 2020) that are associated with certain risks and problematic behaviors (Turel and Serenko, 2012; Matute, 2016; Medrano et al., 2017; Valencia-Ortiz et al., 2021); at the extreme, SN use could result in dependency, addiction, or pathological behaviors (Andreassen et al., 2012; Guertler et al., 2014; Chóliz, 2016a; Echeburúa, 2016; Matute, 2016) that require intervention or the development of preventive programs (Chóliz and Marco, 2012; Marco and Chóliz, 2013; Chóliz, 2016b; Marcos, 2020).

The popularity and massive use of SNs peaked in the second decade of this century, and there is more interest in the problems that could arise from internet and SN excessive use, abuse, problematic use, pathological use, addiction or dependence, and this interest and concern has led scholars around the world to develop numerous studies in the theoretical and applied fields (Echeburúa and Gargallo, 2010; Tao et al., 2010; Turel and Serenko, 2012; Puerta-Cortés and Carbonell, 2013; Salas, 2014; Gil et al., 2015; Matute, 2016; Hussain and Griffiths, 2018; Cabero-Almenara et al., 2020a; Ruiz-Ruano et al., 2020; Lupano and Castro, 2021).

Many researchers have focused on measuring phenomena that do not have clear definitions or understanding of their similarities or differences (excessive use, problematic use, abusive use, addiction, risk of addiction, pathological or compulsive use of the internet and SNs), and related phenomena such as nomophobia (fear of missing out). These scales have been adapted and validated in various languages and parts of the world (Beranuy et al., 2009; Meerkerk et al., 2009; Lam-Figueroa et al., 2011; Andreassen et al., 2012; Karim and Nigar, 2014; Boubeta et al., 2015; Vilca and Vallejos, 2015; Fonsêca et al., 2018a,b; Peris et al., 2018; Sahin, 2018; Vallejos-Flores et al., 2018; da Veiga et al., 2019; Laconi et al., 2019; Cabero-Almenara et al., 2020a,b; García-Umaña and Córdoba, 2020; Santamaria and Vallejos-Flores, 2021). This broad conceptual, instrumental and contextual coverage suggests that interest in this topic has increased worldwide.

The SN Addiction questionnaire (SNA) (Escurra and Salas, 2014) was constructed with the diagnostic and statistical manual - 4 (DSM 4) criteria for the diagnosis of substance addiction. It has 24 items grouped into three dimensions: obsession with SNs, lack of personal control in the use of SNs, and excessive use of SNs (Escurra and Salas, 2014). Its validity has been replicated multiple times, with its factorial structure maintained in some studies (Medrano et al., 2017; Salazar-Concha et al., 2021) and varied in others. Bueno et al. (2019) found five dimensions, and others argue that the instrument is unidimensional (Fonsêca et al., 2018b; Salas-Blas et al., 2020). The latter structure motivated and justified the construction of the SNA-6 (Salas-Blas et al., 2020), but with a different theoretical basis (Griffiths, 2005a), called the *Components Model*, which has six dimensions. Within the SNA-6, each item represents one of these six components. The model of the components of addiction from a biopsychosocial perspective, according to Griffiths (1996, 2005a,b), is applicable to any addictive behavior, with and without substances, and is based on the following components: (a) *Salience*, which consists of the excessive behavior becoming the most important thing for the addicted person, because it dominates his or her thoughts, feelings and behaviors; (b) *Mood change*, which are subjective experiences resulting from the activity, such as ecstasy, numbing and escape from reality; (c) *Tolerance*, in which more time is dedicated to the activity; (d) *Abstinence*, which manifests itself in states of discomfort, trembling, irritability in the absence of the addictive activity, and which are only overcome when the activity is resumed; (e) *Conflicts*, generated between the addict and the people around him/her, such as friends, family, partner, work, school, etc.; (f) *Relapses*, tendency to repeatedly return to previous patterns after experiencing abstinence.

The six items of the brief version of the SNA (SNA-6), come from the original version of 24 items (Escurra and Salas, 2014), they were chosen by expert judges who first located to which of the components each of the 24 items theoretically belonged; once this first classification was done, they were asked to choose two items that they considered theoretically the most representative of each factor. These twelve items were then statistically analyzed to form the six most empirically representative. The items were formulated with the following characteristics: (a) in construct orientation, where the response in the high levels of the scale should correspond to high levels in the construct measured, (b) the use of the instrument is mainly oriented for group description and individual screening, (c) the content shows the temporal intensity of the behavior.

The purpose of this brief version is to facilitate complex research; the development of studies with different populations, both in size and in cultural variety; and studies in which several variables are related, both in predictive research models and in multicausal type. In this type of complex study, the use of short but valid instruments is necessary, as in the present investigation, which focuses on the covariation of SNA with external variables, such as sex, sensation seeking, and symptoms of depression.

Regarding the relationship between SN addiction and sex, there are contradictory results; some studies found that men obtain higher scores (Salas and Escurra, 2014; Müller et al., 2016; Bueno et al., 2019), others found that women have higher scores (Rodriguez and Fernandez, 2014; Peris et al., 2018), and still, others found no differences. A similar situation occurs when comparisons are made by age groups (Beranuy et al., 2009; Salas-Blas, 2019b; Loss et al., 2020); these contradictory data may vary by culture.

Although there is evidence of relationships between sensation seeking and some phenomena, such as risk behaviors, tobacco consumption, alcohol consumption, aggressiveness, antisocial behavior, and substance addictions (Lorca and Sanz, 2003; Nadal, 2008; Merino-Soto and Salas Blas, 2018), there is little research on its relationship with behavioral addictions and specifically, internet and SN addiction; some studies have found positive relationships of impulsivity, lack of control, and gratification seeking (which could also characterize the sensation-seeking) with addictions to the internet and SNs (Koo and Kwon, 2014; Zhang et al., 2015; Becoña, 2016; Clemente et al., 2019; Lupano and Castro, 2021). Based on this evidence and the conviction that other addictions have the same characteristics as substance addictions, that sensation seeking and addiction can be assumed to be positively correlated with SN addiction.

Regarding the relationship between depressive symptoms and addiction to SNs, there are many studies based on applying the first measure of addiction to the internet (Young, 1998), and the results consistently show positive correlations between the two phenomena (Zhang et al., 2015; Becoña, 2016; Błachnio et al., 2016; Hussain and Griffiths, 2018; Peris et al., 2018; Iovu et al., 2020). Therefore, it is expected that this study will find similar results. The study of dependence on technologies, including SNs, is increasing and relevant to many phenomena and activities in people's lives. This work has methodological importance in validating a scale to measure addictions to SN in adolescents and young people in Mexican, across its wide territory and very large population. The effects of irrelevant responses to the construct, such as those *via* the careless/insufficient effort response (C/IE), are addressed. Research has shown such effects to be common (Faust et al., 2012; Meyer et al., 2013), and any such effects must be identified and addressed because they have been shown to affect the classification of subjects (Emons et al., 2007) and the assessment activities for the classification of subjects with possible behavioral dependencies. Additionally, there is an incremental effect of false positives (Faust et al., 2012; Meyer et al., 2013), as well as psychometric properties in general, such as dimensionality, internal structure, and reliability (Liu et al., 2010; Arias et al., 2020). These issues are particularly important and impactful for short scales because the accuracy of respondent classification is linked to the prevalence of careless/insufficient response effort (C/IE) and other factors (Emons et al., 2007). In this regard, there is a call to take C/IE into account in the field of addiction studies (Godinho et al., 2016; King et al., 2018), and the present study addresses this potential problem as the first step toward the main objective of the study.

Therefore, the objective of the present study is to adapt and psychometrically validate the Brief Questionnaire on Addiction to Social Networks (SNA-6) with adolescents and young Mexicans. For this, evidence of the internal structure (i.e., dimensionality, measurement invariance, and reliability; American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014) and associations with other variables were obtained.

Because there is an apparent weakness in the methodology for examining the psychometric performance of items in epidemiological and public health research (Olsen, 1998; Hagquist, 2019), and the SNA can potentially be in these research areas, the present study introduced the study of the quality of response options. Within this framework of methodological rigorousness, it was taken into account that external sources of information (i.e., constructs or external criteria) are drawing attention to give better and more supported conclusions about construct validity (e.g., see Taras and Kline, 2010; Ferrando et al., 2019; Ferrando and Lorenzo-Seva, 2019), so the item analysis included the analysis of association validity with external criterion. This has connections with construct validity at the level of specific items (Taras and Kline, 2010), with content validity (Koller et al., 2017), or with the emergence of differential relationships with external variables (Ferrando and Lorenzo-Seva, 2019).

The following hypotheses served as a framework to guide the interpretation of the results: first, the structure of the response options will show optimal characteristics (i.e., in agreement with Linacre guidelines; see Procedure section); second, the latent structure of the SNA-6 scores will be unidimensional; third, the internal structure properties will be invariant across gender, age, and campus groups; and fourth, the relationships of the SNA-6 with measures of depressive symptoms and sensation-seeking will theoretically converge (different from zero).

MATERIALS AND METHODS

Participants

The population invited to participate comprised Mexican students at the upper secondary level between 15 and 19 years old who were selected to explore potential areas of intervention in subjects related to mental health. The selected population was from five campuses id Cuautla, Cuernavaca, Jiutepec, Temixco, and Tepoztlán in the state of Morelos (Mexico). The selected sample comprised adolescents who met the following inclusion criteria: enrolled as an upper secondary student in one of the five selected schools and providing parental authorization with the informed consent and assent of the participant. Data due to potential careless/insufficient effort responses were excluded.

The initial participating sample included 2,998 individuals (**Table 1**), distributed in the five schools as follows: 934, 371, 818, 347, and 528. For confidentiality, alphabetic campus names (A, B, C, D, and E campuses) were used to mask the true names. After excluding responses apparently due to careless/insufficient effort (see Results section), the effective sample for the study analysis was 2,789. The sex distribution was similar with respect to age (linear $\chi^2 = 3.11$, $p = 0.07$, gamma = 0.04), campus ($\chi^2 = 73.75$, $p < 0.01$, V Cramer = 0.13), marital status ($\chi^2 = 4.62$, $p < 0.05$, Cramer V = 0.05), and semester (linear $\chi^2 = 2.79$, $p = 0.09$, gamma = −0.05). However, there was a slight difference with respect to current employment status ($\chi^2 = 139.09$, $p < 0.01$, Cramer V = 0.223).

Instruments

Because to the natural variability of sampling and in the relations between variables, and to avoid inducing internal structure validity based on previous reports without corroborating it in the actual participant sample (Merino-Soto and Calderón-De la Cruz, 2018; Merino-Soto and Angulo-Ramos, 2020), the internal structure of external construct measures (BSSS and CESD-7) was verified.

TABLE 1 | Distribution of demographic characteristics ($n = 2,789$)[a].

	N	%
Sex		
Female	1406	50.4
Male	1383	49.6
Campus		
A	877	31.4
B	339	12.2
C	760	27.2
D	319	11.4
E	494	17.7
Semester		
2	1067	38.3
4	876	31.4
6	846	30.3
Marital status		
Single	2687	96.3
Married	22	0.8
Unmarried	80	2.9
Work		
Yes	945	33.9
No	1844	66.1
Age		
14	2	0.1
15	567	20.3
16	784	28.1
17	811	29.1
18	495	17.7
19	81	2.9
20	32	1.1
21	9	0.3
Missing	8	0.3

[a]*Sample with exclusion criterion used (C/IE responses).*

Brief Scale of Social Networks Addiction (SNA – 6) (Salas-Blas et al., 2020)

The SNA-6 was constructed from a previous version of 24 items and three dimensions published by Escurra and Salas (2014). From this initial study, the authors found three highly correlated dimensions, which were subsequently reduced in the current version, the SNA-6 (Salas-Blas et al., 2020). The SNA-6 measure evaluates addiction to SNs. The authors performed a content validation of the items using the Griffiths (2005b) component model, which proposes that addictions have six components (salience, mood change, tolerance, abstinence, conflicts, and relapse); in the SNA-6, there is an item that measures one of the factors. The unidimensionality of the SNA-6 was found in Peruvian adolescents (Salas-Blas et al., 2020) and corroborated in the validation in Brazil (Fonsêca et al., 2018b). The items are ordinally scaled in 5 response options (from 1 to 5), from *Not at all* to *Always*. Responses to the items are summed to obtain a total score. The Peruvian version (Salas-Blas et al., 2020) was used in the present study; and to ensure clarity of content, the Mexican co-author of this article verified the appropriateness of its content.

Brief Sensation Seeking Scale (Hoyle et al., 2002)

The BSSS evaluates the attribute of sensation seeking and is generally used to detect the risk of substance use in adolescents. It consists of eight items scaled in five points (from strongly disagree to strongly agree) and produces a unique score. The version for Latino adolescents was used (Stephenson et al., 2007) and was adapted for Peru (Merino-Soto and Salas Blas, 2018). In the present study, one of the Mexican authors of this manuscript checked the content for clarity and conceptual relevance for Mexican adolescents, and both aspects were found to be appropriate. Also, the unidimensionality of the instrument adjusted satisfactorily, weighted least squares means and variance (WLSMV)-$\chi^2 = 436.65$, df $= 20$, $p < 0.01$; CFI $= 0.993$, RMSEA $= 0.034$, SRMR $= 0.048$, WRMR $= 2.69$, with factor loadings between 0.602 y 0.842; the internal consistency was $\omega = 0.87$ (IC95% $= 0.86$, 88, se $= 0.004$).

Center for Epidemiological Studies Depression Scale – 7 (Herrero and Gracia, 2007)

The measurement of depression symptoms (dysphoric mood, poor motivation, concentration, pleasure and sleep) was taken from Santor and Coyne (1997) and adapted to the Mexican context. It consists of seven items oriented toward depression, except item 6 (item 6 needed recoding), scaled in four points (from never to always). For the present study, the one-dimensional model of the CESD-7 was not satisfactory (WLSMV-$\chi^2 = 994.57$, df $= 14$, $p < 0.01$; CFI $= 0.985$, RMSEA $= 0.077$, SRMR $= 0.101$, WRMR $= 4.866$), mainly due to the low load of item 6 (factor load $= 0.067$, se $= 0.02$). Without this item, the fit improved substantially [WLSMV-$\chi^2 = 265.90$, gl $= 15$, $p < 0.01$; CFI $= 0.996$, RMSEA $= 0.027$, SRMR $= 0.055$, WRMR $= 2.839$ (factor loads between 0.440 and 0.935)]. The internal consistency of the six items was $\omega = 0.87$ (95% CI $= 0.86$, 88, se $= 0.004$).

Sociodemographic Information Sheet

A sheet was used to record information about age, sex, semester of study, marital or marital status, and work activity.

Procedure

Data Collection

The questionnaire was administered electronically between April and May 2019. The administration of the survey was directed and supervised by the counselors of each campus, who received training in standardized procedures for administering the survey and resolving situations that may arise. The administration procedure and the order of presentation of the instruments to the participants were the same for each campus: informed consent, demographic questions, and instrument questions. The specific steps were as follows: the consent form was given to the parent or guardian to allow their children to participate in the study, and then, informed consent was provided by each student. Finally, the participants responded collectively in the computer center. The data collection was guided by the principles of the Helsinki Declaration and the Belmont Report in several ways: the anonymity of response, voluntary participation, freedom to withdraw, and confidentiality of information collected. To protect subject anonymity, no identifiable human data were retained. The participants were informed of their right to continue or end their participation at any moment. Finally, the importance of honest responses, careful attention to the instructions and contents of the items, and willingness to resolve doubts about filling out the survey were highlighted.

Ethical Considerations

This study is a part of the research project (HIM/2015/017/SSA.1207; "Effects of mindfulness training on psychological distress and quality of life of the family caregiver") that was approved on December 16, 2014, by the Research, Ethics, and Biosafety Commissions of the Hospital Infantil de México Federico Gómez National Institute of Health in Mexico City. While conducting this study, the ethical rules and considerations for research with humans in Mexico (Sociedad Mexicana de Psicología, 2010) and those outlined by the American Psychological Association (2017) were carefully followed. All family caregivers were informed of the objectives and scope of the research and their rights in accordance with the Declaration of Helsinki (World Medical Association, 2013). The caregivers who agreed to participate in the study signed an informed consent letter. Participation in this study was voluntary and did not involve payment.

Analysis of Data

The analysis was performed in the following sequence: evaluation of response biases, item analysis, internal structure analysis, and relationship with other variables.

Potential Response Biases

The SNA was included in a survey of 11 dimensions measuring psychological, emotional, and attitudinal attributes, with 50 total items, but several of these items nested with other items, for a total of 143 effective items measuring demographic information and other constructs (e.g., sensation seeking). Therefore, this extension of the questionnaire can be associated with the generation of careless/insufficient effort responses (C/IE; Huang et al., 2015; Curran, 2016) especially in adolescents (Cornell et al., 2012; Jia et al., 2018). C/IE can be expressed in response patterns with extreme or little variability (Curran, 2016), and in this sense, and as a general measure of anomalous responses, the presence of multivariate outliers was first verified using distance D^2 (Mahalanobis, 1936), an efficient method to detect possible random responses (Zijlstra et al., 2011). This control reduces the sensitivity of the covariance matrix and the emergence of model specification errors of latent variables due to outliers (Wilcox, 2012).

Because the data were treated categorically in most of the analyses, the detection of outliers was done within a non-parametric approach for non-continuous variables, which are considered discordant observations expressed in *extreme scores* (e.g., O + y G + scores; Zijlstra et al., 2007), which differ from the regular scoring pattern due to their infrequency. G_+ is based on the paired comparison of items to obtain the number of errors according to Guttman (1944). The complementary use of both has been recommended (Zijlstra et al., 2007), but G_+ alone or in combination with the Mahalanobis distance are effective measures for the detection of random responses (Zijlstra et al., 2011, 2013). For the procedures described in this analysis, the R *careless* (Yentes and Wilhelm, 2021) and *mokken* (van der Ark, 2012) programs were used.

Item Analysis

First, descriptive and distributional statistics for the items were obtained. For association analysis, non-parametric indices were chosen to maintain the mainly categorical treatment given to the SNA-6 items. Thus, correlational effect size indices were used (*Glass rank-biserial*, *eta-squared,* and *epsilon-squared*; Tomczak and Tomczak, 2014). Together, these indicators served to investigate the possible differentiation of the items in their relationships with other variables. Quantitatively, this adds support for validity. The R *langtest* (Mizumoto, 2015), *rcompanion* (Mangiafico, 2021), and *MVN* (Korkmaz et al., 2014) programs were applied. Second, to evaluate the structural properties of the SNA response option in detail, item response theory modeling was used.

Properties of Ordinal Scaling

The structural properties of the response options for each SNA item were evaluated using the suggested guidelines of Linacre (1999, 2002a) that the response categories (a) have more than 10 responses, (b) show a monotonic increase in the thresholds with the latent attribute score, (c) have a regular distribution, (d) have ordering thresholds, and (e) have step difficulties that advance at least 1.0 logits but less than 5.0 logits. The application of these properties required evaluating whether the items fit the partial credit model (PCM). Given the relative independence of the mean square statistics [Infit mean square (In-MSQ); outfit mean square (Out-MSQ)] from the sample size, these fit indicators were preferred over the accompanying statistical tests (t test for Outfit and Infit) (Smith et al., 2008). The R eRm program was used (Mair et al., 2021).

Internal Structure

We use the confirmatory factor analysis within the structural equations modeling (CFA-SEM) approach. The WLSMV estimator applied to the polychoric correlations between the items was used since it helps obtain psychometric parameters with greater precision for categorical and non-normal variables (Finney and DiStefano, 2013). In the framework of model comparison, a model with a method factor was included to estimate a possible variance associated with response patterns (Maydeu-Olivares and Steenkamp, 2018). For this, the random intercepts model was implemented (random intercepts factor analysis: RIFA; Maydeu-Olivares and Coffman, 2006). The usual specification for RIFA is to include a factor to all items (i.e., method factor), in addition to the substantive factors (in this study, a single SNA substantive factor), and set their factor loadings to 1 (to establish the equality between them). In the present study, the generalized random intercepts framework was implemented, in which it is specified that the effect of method variance varies in each item (Maydeu-Olivares and Steenkamp, 2018). This contrasts with the presumption of equality of the method variance in the items proposed in the original method (Maydeu-Olivares and Coffman, 2006). The models that presented close fit were those with the following degrees of fit: RMSEA $\leq$ 0.05, SRMR $\leq$ 0.05, and CFI $\geq$ 0.95; Maydeu-Olivares (2017). The reasonable adjustment was recognized when RMSEA $\leq$ 0.08, SRMR $\leq$ 0.08, and CFI $\geq$ 0.90 (Hu and Bentler, 1999). Possible adjustments to improve the model were examined using the statistical power and size of the parameter (i.e., correlated residual) and were implemented by the approach of Saris et al. (2009). Additionally, attention was given to residual correlations greater than.10 (Maydeu-Olivares et al., 2018). SEM was performed with the programs R *lavaan* (Rosseel, 2012) and *semTools* (Jorgensen et al., 2021).

As part of the internal structure study, measurement invariance and reliability were analyzed (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014). Measurement invariance was measured with a bottom-up approach, from an unconstrained model to a strongly constrained model (Stark et al., 2006). Thus, we tested an unconstrained model (configurational invariance) and continued with successive restrictions applied to factor loadings, thresholds (metric invariance), and intercepts (scalar invariance). Given the sample size (> 300; Chen, 2007), the invariance criterion used were CFI < 0.010, SRMR < 0.030, and RMSEA < 0.015 (Chen, 2007). Age and gender were chosen as possible phenotypical sources of measurement variability since both variables are usually involved in studies of invariance in psychosocial measures. The campus of the study was chosen as a context variable where the degree of clustering of shared experiences may be present among students within the campus.

The reliability estimate for the score was made for congeneric measures, with the omega for categorical variables (ω; Green and Yang, 2009). For reference, the alpha coefficient (α) was also estimated. For both coefficients, confidence intervals were created at the 95% level (Bootstrap simulation method; Kelley and Pornprasertmanit, 2016). The R *MBESS* package was used (Kelley, 2019).

Description and Relationship With Other Variables

The description of the direct score of the SNA-6 was made by evaluating its fit to one of the seven distributions of the Pearson system (Ord, 2006). The R *Pearson DS* program (Becker and Klößner, 2021) was used. The relationships of the SNA-6 scores with other variables were examined using linear correlations and multiple line regression controlling for the effects of sex and age. The scores used were the simple sum of the items of the criterion variables (BSSS and CESD-7) and the latent factor of the SNA-6.

RESULTS

Response Biases

The G_+ scores ranged between 0 and 68 ($M = 4.75$. Md = 1, Q1 = 0, Q3 = 6, IQR = 6). For the identification of infrequent cases, cases above the cutoff point based on Tukey's whiskers (Q3 + 1.5 * IQR = 15; Zijlstra et al., 2011, 2013) were chosen. Therefore, 278 participants with $G_+ \geq 15$ had an infrequent response pattern. With respect to D^2, with a cutoff point of 16.81 (gl = 6, $p < 0.01$), 278 cases with a score $D^2 > 16.81$ were identified and removed from the database. D^2 between 16.41 and 73.37 (M = 6, Md = 3.19). The two identification criteria converged on $r = 0.75$ ($p < 0.01$), with an agreement of 209 true positives ($\chi^2 = 2107.1$, gl = 52, $p < 0.01$, *Cramer-V* = 0.83). The effective sample for the following analyses was 2,789 participants (cases removed in total: 209, 6.9%).

Item Analysis

Univariate Description

Table 2 shows the results. The distribution of responses was positive asymmetric, with a higher density around the infrequent expression of behavior dependent on SNs. The response difference in the items was statistically significant (Friedman - $\chi^2 = 485.32$, $p < 0.01$) but trivial in size (*Kendall-W* = 0.035). The association with sex was approximately zero and not statistically significant; in the same way, in the associations with age, semester, and school, the exact variation of the associations was between 0.0001 and 0.003, none with statistical significance. Multivariate normality was rejected (Henze-Zirkler β = 382.3929, $p < 0.01$).

Structure of the Response Options

The results are presented in **Table 3**. The χ^2 adjustment statistic to the PCM indicated that items 2, 3, 4, and 6 work appropriately with this model ($p > 0.10$); items 1 and 5 did not ($p < 0.01$). The Out-MSQ and In-MSQ values for each item were similar, greater than 0.50 and less than 1.30. According to what is suggested in this type of modeling (Wright and Linacre, 1994; Linacre, 2002a), based on Out-MSQ and In-MSQ, the items can be considered in a productive measurement range. The MSQ values in items 1 and 5 were also in an acceptable range. The discrimination of the items was high (> 0.49; between 0.50 and 0.82), and they were moderately similar. Additional and global indicators of fit of the model were separation reliability = 0.76, observed variance (squared standard deviation) = 2.48, and mean square measurement error (model error variance) = 0.58. Taken together, of all the results presented in this paragraph, the PCM model seems to be sufficient to parametrically model the structure of the response options.

Frequency of Categories

All response options met the criteria ($n \geq 10$; **Table 2**). Particularly, the response option where the lowest distributional density occurred (*always* option) was between 4 and 6.4 times greater than the criterion $n \geq 10$. The prevalence of response ranged from 1.3 to 6.4% for each item. Complementarily, the possible floor and ceiling effects were observed, defined as the high percentage in the highest or lowest response option, respectively. Considering the frequencies of the options (**Table 2**), the first two response options exceed 15% prevalence, and the first option had a very high prevalence (% > 55), indicating a floor effect in each item of the SNA.

Monotonic Increase in Response Options

In **Figure 1** – 1.1, the thresholds show a monotonic and regular advance (horizontal axis) in the range −3 to + 4 of the latent attribute (vertical axis). This regularity is linear and is slightly

TABLE 2 | Descriptive properties for the social networks addiction (SNA)-6 items.

	Correlations									
	Sna1	**Sna2**	**Sna3**	**Sna4**	**Sna5**	**Sna6**	**Sex**	**Age**	**Sem**	**Campus**
Sna1	1						0.00	0.00	0.00	0.00
Sna2	0.59	1					−0.03	0.00	0.00	0.00
Sna3	0.58	0.76	1				0.00	0.00	0.00	0.00
Sna4	0.56	0.68	0.73	1			−0.03	0.00	0.00	0.00
Sna5	0.51	0.57	0.61	0.62	1		−0.03	0.00	0.00	0.00
Sna6	0.59	0.69	0.72	0.74	0.65	1	−0.06	0.00	0.00	0.00

	Frequencies of options					
	Sna1	**Sna2**	**Sna3**	**Sna4**	**Sna5**	**Sna6**
Op 1	1689	1902	1828	1871	1924	2217
Op 2	838	724	732	677	533	494
Op 3	336	257	295	295	299	178
Op 4	88	75	79	94	140	69
Op 5	47	40	64	61	102	40

	Statistic descriptive				
	M	**SD**	**Sk**	**Ku**	**CvM**
Sna1	1.61	0.84	1.43	1.93	54.98
Sna2	1.48	0.78	1.77	3.24	75.60
Sna3	1.53	0.82	1.69	2.84	68.67
Sna4	1.52	0.82	1.70	2.86	72.14
Sna5	1.54	0.90	1.70	2.23	81.43
Sna6	1.34	0.71	2.45	6.64	112.05

All associations were statistically non-significant. Sna: items from the social network addiction scale (SNA-6). Op: response options: never (op1), rarely (op2), sometimes (op3), almost always (op4), always (op5). Sk: skewness coefficient. Ku: kurtosis coefficient. CvM: Cramer von Mises normality test. Sem: semester.

TABLE 3 | Partial Credit model Fit for ítems of SNA-6.

	PC Fit results(Partial Credit model)					Parameter items from PC model					
	χ^2(df: 1757)	**Out-MSQ**	**Out-t**	**In-MSQ**	**In-t**	**Disc.**	**Loc**	**Thr_1**	**Thr_2**	**Thr_3**	**Thr_4**
Sna1	2250.57*	1.28	7.84	1.24	6.92	0.50	0.71	−2.51	−0.38	1.87	3.86
Sna2	1435.51	0.81	−5.18	0.79	−6.38	0.73	1.24	−1.89	0.07	2.16	4.64
Sna3	1276.95	0.72	−8.30	0.69	−9.98	0.79	0.83	−2.05	−0.21	1.99	3.62
Sna4	1375.67	0.78	−6.22	0.75	−7.68	0.75	0.92	−1.94	−0.25	2.12	3.77
Sna5	1900.47*	1.08	1.77	1.11	3.14	0.60	0.64	−1.59	−0.67	0.91	3.93
Sna6	963.96	0.54	−9.92	0.63	−10.57	0.82	1.49	−0.99	0.43	2.46	4.09

*Sna: items from the social network addiction scale (SNA-6). Thr_i: threshold parameter. Loc: location parameter. Disc.: discrimination parameter. In-t, Out-t: statistical test for In-MSQ and Out-MSQ. df: degree free. *p < 0.05.*

altered in items 4 and 5, but they do not seem to suggest a substantial departure from the observed linear increase. In conclusion, this structural criterion was met.

Regular Observation Distribution

In **Figure 1** – 1.2, the distributional characteristics are shown by kernel densities (Gaussian smoothing, adjusted = 3). Similar characteristics are shown in all the items in the trend toward unimodality of the responses, in the prevalence of the first two adjacent categories, in the long tails to the left, and in the generally strong distributional similarity in all the items. According to these results, the item distributions were regular.

Ordering Thresholds (Step Calibrations Advance)

In **Figure 2**, the ordering of the k – 1 threshold (k = number of response options) in two moments of the data is presented: before (**Figure 2**.1) and after (**Figure 2**.2) the removal of potentially biased data. A strong inconsistency occurred in the data with potential responses with insufficient effort, in the first two thresholds, and the last two thresholds. This inconsistency of the first two options involved variable and wide spacings, while the inconsistency in the last two thresholds was small. The disorder is also observed in item 3, and little distinction is observed in the last two thresholds of item 4. In contrast to the above, in the lower subgraph (**Figure 2**.2), consistency was observed in the ordering

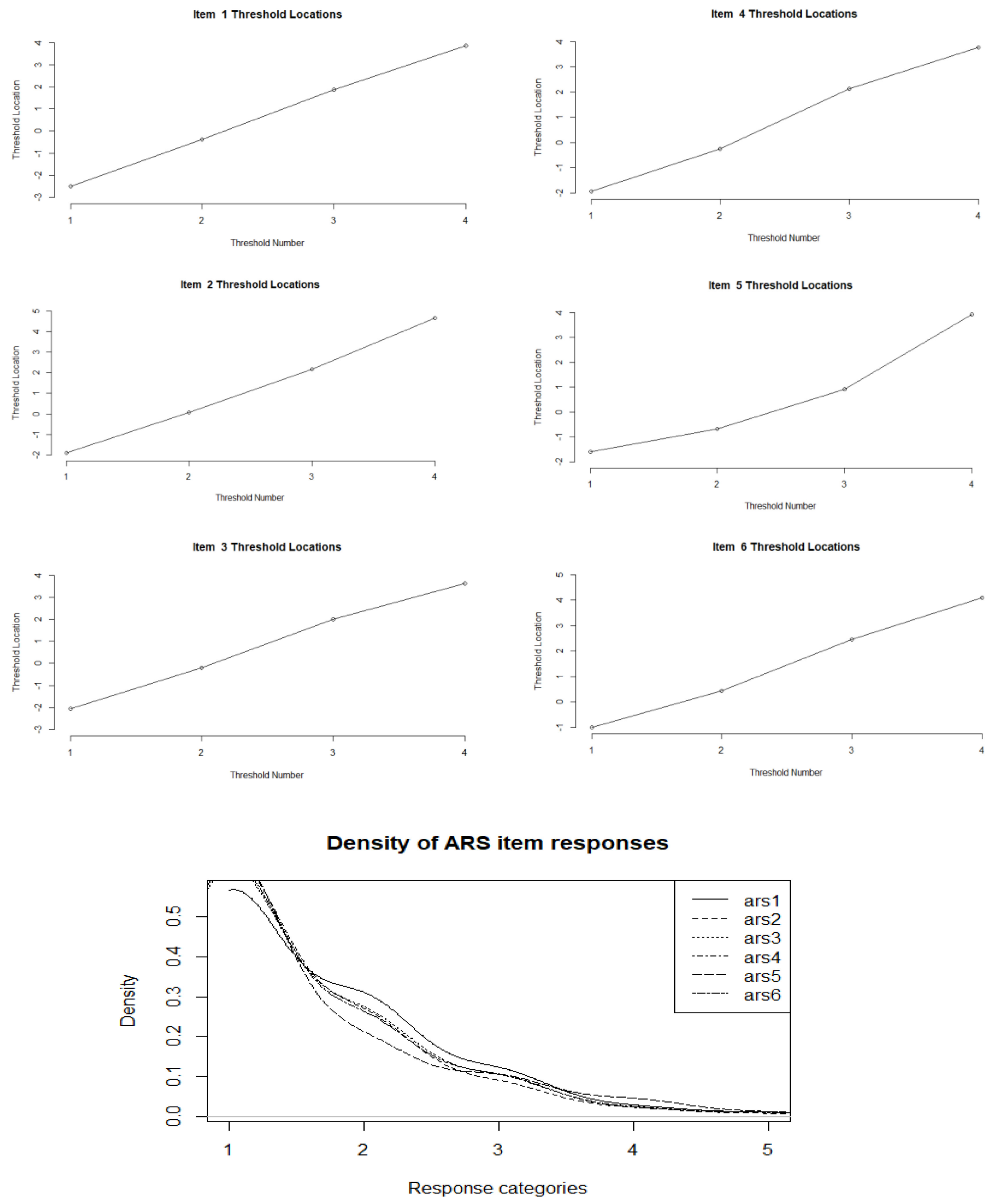

FIGURE 1 | Evaluation of threshold monotonicity and distributional regularity (SNA-6 scale).

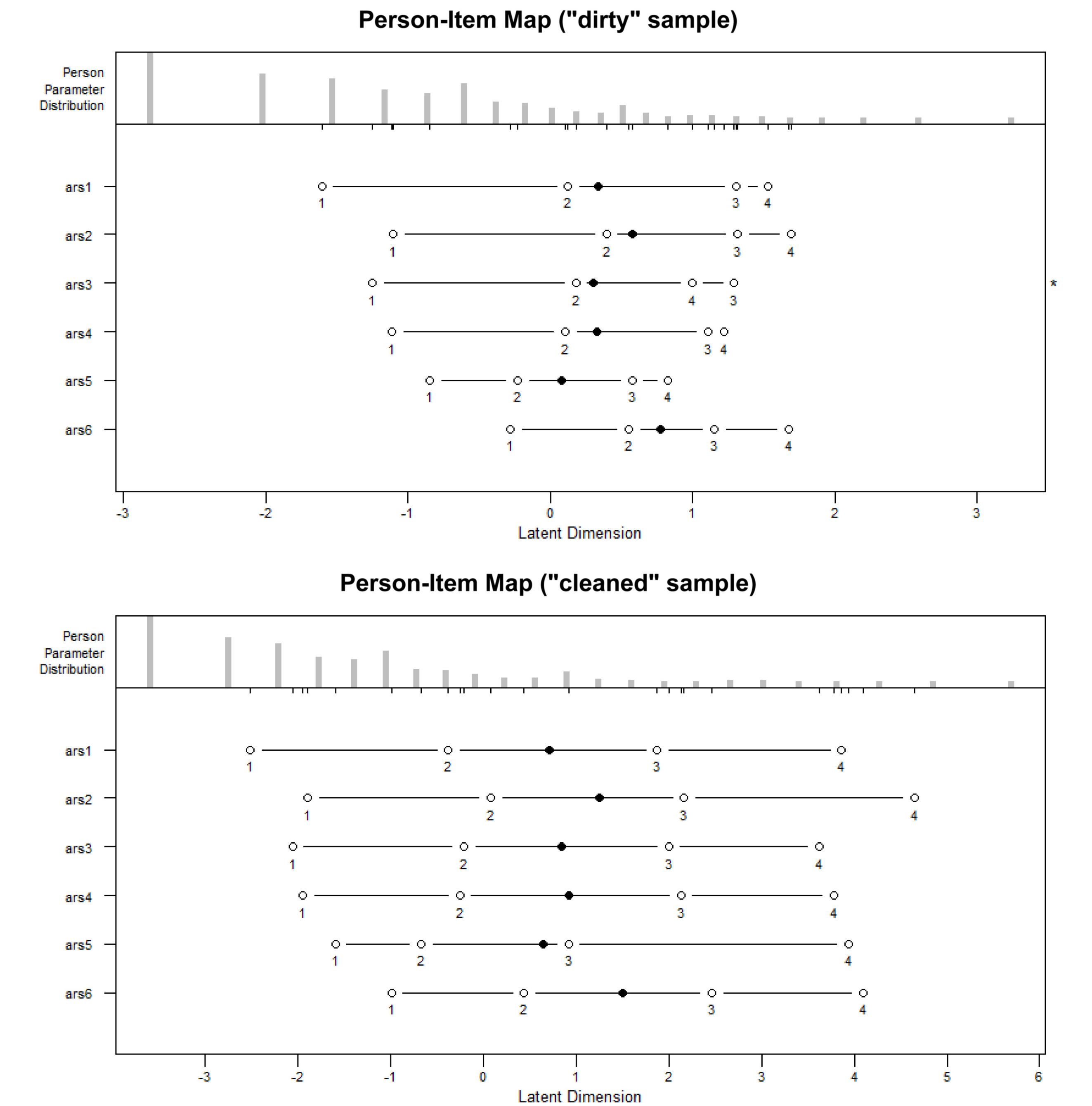

FIGURE 2 | Orderings thresholds: clean and "dirty" sample (SNA-6 scale).

of the thresholds in an order expected for the structure of the items. In conclusion, this criterion was met.

Step Difficulties Advance: At Least 1.0 Logits, but Less Than 5.0 Logits

The average difference, deduced from the thresholds presented in **Table 3** (Thr_1 to Thr_4), between each threshold was, from 1 to 4, −1.66 (min = −2.13, max = −0.92), −2.09 (min = −2.37, max = −1.5) and −2.06 (min = −3.02, max = −1.62), respectively. All these values are below the maximum (< 5) and above the minimum cutoffs (> 1). Thresholds 1 and 2 of item 5 were the exceptions, but this discrepancy with the criteria can be considered insubstantial for practical purposes; therefore the criterion was considered to be fulfilled. Altogether, this feature can be considered complete. According to the results in this section, the hypothesis on the structure of response options is fulfilled.

Descriptive and Association With External Variables Results

Internal Structure

Dimensionality

The one-dimensional model was rejected due to its statistical significance (null model: substantive or theoretical model), but the approximate fit indices were satisfactory because they met the criteria (**Table 4**; one-dimensional model heading). The WRMR was slightly higher than the criterion (WRMR > 1.0). Factor loadings tended to be greater than 0.80 (except item 1). On the other hand, in the estimation of the model with the method factor (F_{met}), the first adjustment required setting a parameter to achieve adequate identification; thus, item 2, which obtained a very low load on F_{met} in the method factor model, was set at 0.0. With this added specification, the fit improved in all indices (i.e., random intercepts, RI; **Table 4**, heading one-dimensional model with method factor), and they were superior to those in the model without the method factor. The RI factor loads were not equal and were below.45 (between 0.168 and 0.421). The percentage of variance at the item level, attributed to the RI factor, varied between 2.8% (item 3) and 17.7% (item 6), and at the factor level, approximately 4.5% of the variance in the substantive factor was due to method variance, an amount that can be considered low. Controlling for the effect of the method, the factorial loads of the substantive factor remained high and were above.70. On the other hand, the modification indices (Md = 3.80, min = 0.015, max = 35.63) were relatively small, and although some were statistically significant, the standardized correlated residuals ranged between −0.05 and 0.07 (Md = −0.001) and were trivial in size. Therefore, no specification based on modifying indices was needed. The instrument dimensionality hypothesis (one dimension) is accepted.

TABLE 4 | Unidimensional and unidimensional with method factor models.

	Unidimensional	Unidimensional + factor method	
	F	F	Fmet
Sna1	0.750	0.717	0.217
Sna2	0.895	0.945	0.000 (fixed)
Sna3	0.920	0.906	0.168
Sna4	0.902	0.843	0.321
Sna5	0.822	0.741	0.392
Sna6	0.938	0.857	0.421
Var	0.563	1 (fixed)	0.047
Reliability	0.91	0.825	0.086
WLSV-χ^2(df)	53.540 (9)**	6.927 (4)	
CFI	0.999	1.000	
RMSEA(90% CI)	0.042 (0.032,0.053)	0.016 (0.00,0.036)	
SRMR	0.018	0.008	
WRMR	1.17	0.421	

*Sna: items from social network addiction scale (SNA-6). F: factor loadings for unidimensional factor. Var: Factor Variance. Fmet: Method factor (random intercepts). Reliability: omega coefficient. **p < 0.01.*

Measurement Invariance

To measure invariance across the age of the participants, the groups were recategorized to compare balanced groups of 14- to 16-year olds (n = 1,353) and 17- to 21-year olds (n = 1,428). In **Table 5**, the results indicate equivalence across sex, age, and campus groups. It was maintained up to the scalar invariance level (this is, equality of intercepts). As a consequence of the above, the hypothesis of the invariant properties of internal structure is accepted.

Reliability

The coefficients ω (0.91; 95% CI = 0.90,0.92; se = 0.004) and α (0.91; 95% CI = 0.90,0.92; se = 0.004) of the scores were nearly equal. The standard error of measurement in the total sample was 1.23 (SD = 4.10); in men and women, it was 1.32 (SD = 4.43) and 1.12 (SD = 3.74), respectively.

Association With Other Variables, Descriptive, and Difference Results

The distribution of the SNA-6 score was apparently not unimodal (D test = 0.07, p < 0.001). The parametric distribution with which it seems to fit was Pearson Type II (i.e., beta symmetric), since the model selection criteria were more minimized (log likelihood = −2.05; Akaike Information Criterion, AIC = 10.11, Bayes Information Criterion, BIC = 8.27) than the other distributions of the Pearson system, from 0 to VII (log likelihood < −3.36; AIC < −7.47; BIC > 12.28). The parameters for this distribution are shape (skewness) = 0.67, location = 2.05, and scale = 7.25.

On the other hand, age (B = −0.19; β = −0.05, 95% CI = −0.18,0.07; p = 0.003) and sex (B = 0.44; β = 0.15, 95% CI% = −0.25,0.36; p = 0.004; men coded as “2”) produced variability on the latent variable of SNA-6 but a size that can be considered weak and moderate, respectively (β < 0.20: weak, β < 0.50 moderate; Acock, 2014, p. 272). Regarding the campus, only campus 241 was statistically significant, and the variability can be considered weak (B = 0.47; β = 0.05, 95% CI = −0.34,0.45; p = 0.01).

In obtaining evidence of validity with other constructs, the latent linear correlations of the SNA-6 score with sensation seeking and depressive symptoms (BSSS and CESD-7, respectively) are shown in **Table 6**. The size of these latent correlations was small, and they were essentially unchanged after controlling for the effects of sex and age. All correlations were positive; accordingly, the hypothesis of association with depressive symptoms and sensation-seeking is accepted.

DISCUSSION

The psychometric properties of the SNA-6 (Salas-Blas et al., 2020) were studied in the Mexican context, a measure of addiction to SNs created in Peru with potential psychometric characteristics that allow its scores to be interpreted from an emic framework. The study focused on the content of the items, the internal structure, and the relationship with external variables. As an additional note with methodological implications, the

TABLE 5 | Measurement invariance results for the SNA-6 scale.

	WLSMV χ^2	CFI	RMSEA (CI 90%)	SRMR	Differences		
					Δ_{CFI}	Δ_{RMSEA}	Δ_{SRMR}
Sex							
Configurational(df = 18)	56.364	1.00	0.039 (0.028,0.051)	0.019	–	–	–
Thresholds + loadings(df = 35)	104.61	0.999	0.038 (0.030,0.046)	0.022	0.001	0.001	−0.003
Intercepts(df = 40)	118.548	0.999	0.038 (0.030,0.045)	0.022	0.00	0.00	0.00
Age							
Configurational(df = 18)	59.22	0.999	0.041 (0.029,0.052)	0.019	–	–	–
Thresholds + loadings(df = 35)	87.85	0.999	0.033 (0.024,0.042)	0.023	0.00	0.008	−0.004
Intercepts(df = 40)	91.66	0.999	0.030 (0.02,0.039)	0.023	0.00	0.003	0.000
Campus							
Configurational(df = 18)	68.17	1.00	0.030 (0.014,0.044)	0.020	–	–	–
Thresholds + loadings(df = 35)	132.15	1.00	0.017 (0.00,0.029)	0.024	0.00	0.013	−0.004
Intercepts(df = 40)	152.46	1.00	0.016 (0.00,0.027)	0.024	0.00	0.001	0.000

analysis performed began with the identification of participants who likely provided C/IE responses. The prevalence obtained from C/IE was 6.9%, which is not far from the wide range of prevalence reported in other studies (Arias et al., 2020). Although it seems small with respect to the participant sample size (initially, $n = 2{,}998$), the literature has consistently shown the existence of this C/IE response pattern in measures based on self-report surveys and its consequences on psychometric and non-psychometric results (Zijlstra et al., 2007, 2011, 2013; Meade and Craig, 2012; Meyer et al., 2013; King et al., 2018; Arias et al., 2020). Even with 2.5% random responses, the relevant estimates for the psychometric interpretation of the scores are inflated, something that particularly occurs in highly skewed and low prevalence samples (Cornell et al., 2012; Huang et al., 2015; Jia et al., 2018; King et al., 2018), as occurs in the distribution of SNA-6 scores. A practical implication of these results on the quality of the SNA-6 scores is that adjustments to the reliability estimates may be required with available methods (e.g., Fong et al., 2010), but the detection and removal of cases may be a more common, parsimonious, and secure solution when the data are in the hands of the user or researcher.

In the analysis of the SNA-6 response categorization, two things can be highlighted. First, this structure of response options worked appropriately in the present sample because none of the quality indicators proposed by Linacre (1999, 2002a) were challenged. It is also true that some of the items and thresholds did not meet these exact criteria. However, these discrepant values were not severely distanced from the criteria, and for practical purposes, they can still be considered within the chosen criteria. Even with these optimal results, some issues seem to require attention. The first of these is the floor effect found for all items, expressed in very high values (approximately 50% response in the first category, the lowest). As a consequence, the last two answer options were infrequently chosen. In health status measures, the criterion is generally > 15% to identify a significant floor or ceiling effect (Terwee et al., 2007), but in psychosocial measures, there is not a consensus or disseminated criterion in the scientific community. Therefore, the severity or rationality of the floor effect found must be evaluated in relation to the construct and expected use as defined in the present study. In this sense, the SNA measures a construct characterized by the greater intensity of the attribute and linked to inappropriate or maladaptive behaviors; since the participating sample was chosen to represent the distribution of this characteristic in the general community, a strong distributional asymmetry is expected, with greater density in the low scores. As corroborated in the descriptive and distributional analyses at the item level, what happens is precise that the responses of excessive dependence to SNs are not intense but rather low dependence. In summary, the potential floor effect problem is associated with the distributional asymmetry of the scores (Koedel and Betts, 2010), and it is not a constructed problem. Usually, attention is given to problems resulting from the floor effect, produced by the asymmetry of the score distributions (Koedel and Betts, 2010), such as the modeling of group differences (Šimkovic and Träuble, 2019). However, with modern robust analysis methods, this is not always a problem, especially when models such as generalized linear models are used (Šimkovic and Träuble, 2019). Moreover, the scores with asymmetric distribution of the SNA can be modeled with gamma, beta-binomial, and other distributions (Šimkovic and Träuble, 2019).

The second aspect that we highlight in the analysis of the SNA-6 response categorization is that although the results of the SNA-6 response options structure were adequate with the Linacre (1999, 2002a) criteria, the user can still decide whether the response scaling should be optimized, for example, toward a scaling with fewer response options due to the floor effect. In this situation, three options seem to be available to make modifications without losing the ordinal nature of the response and maintain the ordering of the thresholds (Zhu et al., 1997). There is evidence that indicates that the results of optimizing the response categorization, made from a Rasch approach (as applied in the present study), can be stable and reproducible in similar samples (Zhu, 2002). However, since the five SNA response categories remained optimal, the results should be replicable in similar samples.

TABLE 6 | Latent correlations and descriptive statistics for scores.

	SNA – 6	BSSS	CESD-7
Correlations			
SNA-6	1	0.24**	0.27**
BSSS	0.25**	1	0.25**
CESD-7	0.25**	0.25**	1
Descriptives			
M	9.05	24.88	13.25
SD	4.1	8.49	4.61
Sw	2.06	−0.28	0.22
Ku	5.19	−0.71	−0.54

Below the diagonal: latent correlations, not controlling for sex and age. Above the diagonal: latent correlations, controlling for sex and age. SNA-6: social network addition score. BSSS: sensation seeking score. CESD-7: depressive symptom score. Sw and Ku: skew and kurtosis coefficients.

As an additional note regarding the evaluation of the structure of the response options of the SNA, although the predominant tendency of the items was to be below 1.0 in MSQ, it did not decrease the quality of the attribute measurement. Because the instrument was constructed with techniques derived from the classical theory test, high response consistency is expected and used to select the items in the study by the SNA authors (Salas-Blas et al., 2020). Within the Rasch framework, this usually points to redundancy in responses and a highly predictable pattern of responses; however, this characteristic is not always a problem (Wright and Linacre, 1994; Linacre, 2002b).

The item-level analyses also included the association with external variables, to provide utility for construct interpretation at the level of specific behaviors (i.e., items; Koller et al., 2017), in relation to testing fairness and the possible emergence of differential item functioning (DIF), by showing possible context dependencies or phenotypic traits (e.g., sex and age), as well as for psychometric hypothesis formation at the level of scores and item selection. Here, the association between the SNA-6 items and the chosen variables was essentially zero (sex, age, study semester, and campus), and suggests that the specific contents do not covary in a magnitude that may indicate a differential functioning or multilevel approach to understanding social network addiction in the Mexican sample.

The SNA was unidimensional and showed high similarity in its factorial loadings. This replicates the result of the original study in the Peruvian sample (Salas-Blas et al., 2020). Factor loadings significantly exceeded.65, and the fit was satisfactory at a close high fit level (Maydeu-Olivares, 2017) and much more satisfactory for the random intercept model (RIFA model). Because the RIFA showed an excellent fit, one may wonder exactly the intrinsic mechanism of this RI factor. The RIFA model captures a wide range of method effects (Podsakoff et al., 2003; Steenkamp and Maydeu-Olivares, 2021), and the specific origin of this source of variability cannot be specified without an analysis of the characteristics of the instrument and participants examined in the sample. Due to the high fit obtained from the one-dimensional model without including RI (RMSEA $\leq$ 0.05, SRMR $\leq$ 0.05, and CFI $\geq$ 0.95; Maydeu-Olivares, 2017), the interpretation of the SNA scores can exclude the RI factor (Steenkamp and Maydeu-Olivares, 2021), and the one-dimensional model without RI can be accepted as a reasonable representation of the SNA in the Mexican sample.

The apparent optimal quality of the SNA in the Mexican sample was also enhanced by the significant comparisons between groups that can be made, specifically between groups based on sex, age, and study establishment. Indeed, the SNA can give equivalent results for its structural properties because the evaluation of the measurement invariance was satisfactory. Because both biological attributes seem to be common variables investigated to examine the variability of dependency behaviors, comparisons according to sex and age would not be biased due to the structural properties of SNA-6 in the Mexican sample. This result was similar to that of the original SNA-6 study.

Regarding the measurement invariance, the levels reached suggest that the attribute can be measured equivalently between the groups compared. This constancy of the evaluated parameters (i.e., factorial faces, dimensionality, and intercepts) reduces the interpretation of the group differences toward the absence of this equivalence or the differential functioning of the items.

Regarding the distributional form, the density function that can best describe the SNA-6 score by population is Pearson Type II, a special case of the beta distribution. This means that the behavior measured by the SNA-6 shows a higher density in the lower areas of the scores, that is, with lower intensity of the attribute. Due to the nature of the construct measured by the SNA-6, a statistically normal distribution cannot be assumed, but a strongly asymmetric one can be. This does not represent a problem for the theoretical understanding of the construct in the population, since the normal distribution is realistically unlikely (Joo et al., 2017), and precision is required to describe characteristics with high distributional skewness (Trafimow et al., 2019). A practical implication is that the generation of scales must account for not only the mean and standard deviation but also, and at least, distributional skewness (Trafimow et al., 2019). Additionally, a floor effect of the scores is likely (Koedel and Betts, 2010), as discussed above.

Statistically significant differences were detected when comparing the data by age (lower ages scored slightly more than the higher ages), which partially confirms what was found by Salas-Blas (2019a). Likewise, the data are different when compared by sex (men have higher scores than women), which confirms the findings of Salas and Escurra (2014) and Müller et al. (2016). In both results obtained, it is important to highlight that the variability of the SNA-6 due to sex and age can be considered small and moderate, respectively, an issue that is quite frequent when psychological variables are related to biological variables.

It should be noted that these contradictory data on trends in behavioral addictions by age and by sex may be related to sociodemographic variables such as the characteristics of the city in which the participants reside, access to the internet, the place where they connect (home or public cabins), the technological facilities they have access to and some issues related to family functioning; a possible more systematic analysis of the studies carried out could allow us to determine which sociocultural variables could be conditioning one or the other result.

In the evidence of the relationships of SNA with other constructs, weak positive linear associations were found with the attribute of seeking sensations and symptoms of depression, and these were in a positive direction, a fact that confirms what was theoretically proposed. This size of the association seems to be common because with the sensation-seeking score (BSSS), the associations of constructs and behavioral criteria tend to be small (Hoyle et al., 2002; Lorca and Sanz, 2003; Stephenson et al., 2007) and therefore weak but statistically significant. Some studies that found these links do not contain the calculation of the strength or magnitude of the effect, so it is difficult to make comparisons. The relationship of these variables with the SNA-6 probably sets a framework of the expected size, since this phenomenon is observed in many correlational studies in psychology in which the magnitude of the relationship is low but statistically significant. (a phenomenon that could be important to study more carefully in the future). Together, these associations support the interpretation of the SNA-6 scores regarding their theoretical link with sensation seeking and depression symptoms.

Regarding the limitations of the study, no answers were obtained on the intensity of the use of SNs, and therefore, the link of this variable with dependent behavior could not be evaluated. Likewise, the representativeness of the population is not guaranteed, since the sample came from a single Mexican state, and therefore, a generalization cannot be made toward the total Mexican population. Studying the variables with self-report instruments can generate reasonable doubts about the results.

CONCLUSION

The present study provides strong evidence of the validity and reliability of the SNA-6 questionnaire (Salas-Blas et al., 2020) in the Mexican context. These satisfactory properties included (a) an adequate structure of the response options, with potential for improvement by reducing the number of response options; (b) a replicable unidimensionality of the scores with respect to the study in Peruvian adolescents; (c) appropriate reliability values for screening assessments; (d) coherent theoretical relationships with measures of depressive symptoms and sensation seeking. Although the study was conducted in a representative sample of Mexico, the results obtained can be taken as reference values for contrast and support possible conclusions about the replicability of the present results. Given the convergence with studies in other countries (e.g., Peru and Brazil), the SNA-6 can be a useful tool to investigate SN addiction in Latin American countries in general and Mexico in particular.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The present research was approved by the Commissions of Research, Ethics and Biosafety (Comisiones de Investigación, Ética y Bioseguridad), Hospital Infantil de México Federico Gómez National Institute of Health. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

CM-S designed the study and performed the statistical analyses. BP-A acquired and validated the data. ES-B and BP-A contributed to the interpretation of the results. CM-S, ES-B, and BP-A approved it for publication. FT-T contributed to the writing, review, and editing and funding acquisition. CM-S and FT-T contributed to the visualization, project administration, and supervision. All authors drafted the initial and final version of the manuscript.

FUNDING

This work was one of the results of the research project *HIM/2015/017/SSA.1207, "Effects of mindfulness training on psychological distress and quality of life of the family caregiver."* Main researcher: Filiberto Toledano-Toledano. The present research was funded by federal funds for health research and was approved by the Commissions of Research, Ethics and Biosafety (Comisiones de Investigación, Ética y Bioseguridad), Hospital Infantil de México Federico Gómez National Institute of Health. The source of federal funds did not influence the study design, data collection, analysis, interpretation or decisions regarding publication.

ACKNOWLEDGMENTS

We thank Miguel Ángel Núñez Benítez for his support.

REFERENCES

Acock, A. C. (2014). *A Gentle Introduction to Stata*. Texas: Stata Press.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

American Psychological Association (2017). *Ethical Principles of Psychologists and Code of Conduct. With the 2016 Amendment to Standard 3.04*. Washington, DC: American Psychological Association Press.

Andreassen, C. S., Torsheim, T., Brunborg, G. S., and Pallesen, S. (2012). Development of a Facebook addiction scale. *Psychol. Rep.* 110, 501–517. doi: 10.2466/02.09.18.pr0.110.2.501-517

Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., and Arias, B. (2020). A little garbage in, lots of garbage out: assessing the impact of careless responding in personality survey data. *Behav. Res. Methods* 52, 2489–2505. doi: 10.3758/s13428-020-01401-8

Becker, M., and Klößner, S. (2021). *PearsonDS: pearson distribution system. R package version 1.2*. Vienna: R Core Team.

Becoña, E. (2016). “Factores de riesgo y de protección en el uso problemático de internet [Risk and protection factors in problematic internet use],” in *Abuso de Internet: Antesala Para la Adicción al Juego de Azar Online*, ed. E. Echeburúa (Madrid: Pirámide), 35–60.

Beranuy, M., Chamarro, A., Graner, C., and Carbonell, X. (2009). Validación de dos escalas breves para evaluar la adicción a internet [Validation of two short scales to assess internet addiction and mobile abuse]. *Psicothema* 3, 480–485.

Błachnio, A., Przepiorka, A., and Rudnicka, P. (2016). Narcissism and self-esteem as predictors of dimensions of Facebook use. *Pers. Individ. Differ.* 90, 296–301. doi: 10.1016/j.paid.2015.11.018

Boubeta, A. R., Salgado, P. G., Folgar, M. I., Gallego, M. A., and Mallou, J. V. (2015). PIUS-a: problematic internet use scale in adolescents. Development and psychometric validation. *Adicciones* 27, 47–63. doi: 10.20882/adicciones.193

Bueno, R., Portillo, A., and Barboza Navarro, E. (2019). Indicadores de adicción a las redes sociales y factores de personalidad eficaz en escolares de secundaria de Lima [Indicators of social media addiction and efficient personality factors in high school students in Lima]. *Rev. Psicol. Hered.* 12, 12–22. doi: 10.20453/rph.v12i2.3642

Cabero-Almenara, J., Martínez-Pérez, S., Valencia-Ortíz, R., Leiva-Núñez, J. P., Orellana-Hernández, M. L., and Harvey-López, I. (2020a). La adicción de los estudiantes a las redes sociales on-line: un estudio en el contexto latinoamericano [The addiction of students to online social networks: a study in the Latin American Context]. *Rev. Complut. Educ.* 24, 1–11. doi: 10.5209/rced.61722

Cabero-Almenara, J., Pérez-Diez, J. L., and Valencia-Ortíz, R. (2020b). Escala para medir la adicción de estudiantes a las redes sociales [Scale to measure the students' addiction to online social media]. *Converg. Rev. Cienc. Soc.* 27, 1–29. doi: 10.29101/crcs.v27i0.11834

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Model. Multidiscip. J.* 14, 464–504. doi: 10.1080/10705510701301834

Chóliz, M. (2016a). “Adicción a redes sociales: conceptualización del problema, evaluación y prevención [Social media addiction: conceptualization of the problem, evaluation an prevention],” in *Abuso de internet: antesala para la adicción al juego de azar online*, ed. E. En Echeburúa (Madrid: Pirámide), 79–100.

Chóliz, M. (2016b). Prevención de las adicciones tecnológicas en la adolescencia [Prevention of technological addictions in adolescence]. *Padres Maest.* 389, 53–59. doi: 10.14422/pym.i369.y2017.008

Chóliz, M., and Marco, C. (2012). *Adicción in a Internet y Redes Sociales. Tratamiento Psicológico [Internet and Social Media Addiction. Psychological Treatment]*. Madrid: Alianza Editorial.

Clemente, L. A., Guzmán, I., and Salas-Blas, E. (2019). Adicción a redes sociales e impulsividad en universitarios de Cusco [Social networks addiction and impulsivity in university students from Cusco]. *Rev. Psicol.* 8, 13–37.

Cornell, D., Klein, J., Konold, T., and Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychol. Assess.* 24, 21–35. doi: 10.1037/a0024824

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *J. Exp. Soc. Psychol.* 66, 4–19. doi: 10.1016/j.jesp.2015.07.006

da Veiga, G. F., Sotero, L., Pontes, H. M., Cunha, D., Portugal, A., and Relvas, A. P. (2019). Emerging adults and Facebook use: the validation of the Bergen Facebook Addiction Scale (BFAS). *Int. J. Ment. Health Addict.* 17, 279–294. doi: 10.1007/s11469-018-0018-2

Echeburúa, E. (2016). *Abuso de Internet: Antesala Para la Adicción al Juego de Azar Online*. Madrid: Pirámide.

Echeburúa, E., and Gargallo, P. (2010). Adicción a las nuevas tecnologías y a las redes sociales en jóvenes: un nuevo reto [Addiction to new technologies and to online social networking in young people: a new challenge]. *Adicciones* 22, 91–96. doi: 10.20882/adicciones.196

Emons, W. H., Sijtsma, K., and Meijer, R. R. (2007). On the consistency of individual classification using short scales. *Psychol. Methods* 12, 105–120. doi: 10.1037/1082-989x.12.1.105

Escurra, M., and Salas, E. (2014). Construcción y validación del cuestionario de adicción a redes sociales (ARS) [Construction and validation of the questionnaire of social networking addiction (ARS)]. *Liberabit* 20, 73–91.

Faust, K. A., Faust, D., Baker, A. M., and Meyer, J. F. (2012). Refining video game use questionnaires for research and clinical application: detection of problematic response sets. *Int. J. Ment. Health Addict.* 10, 936–947. doi: 10.1007/s11469-012-9390-5

Ferrando, P. J., and Lorenzo-Seva, U. (2019). An external validity approach for assessing essential unidimensionality in correlated-factor models. *Educ. Psychol. Meas.* 79, 437–461. doi: 10.1177/0013164418824755

Ferrando, P. J., Lorenzo-Seva, U., and Navarro-Gonzalez, D. (2019). unival: An FA-based R package for assessing essential unidimensionality using external validity information. *R J.*, 1–10.

Finney, S. J., and DiStefano, C. (2013). “Nonnormal and categorical data in structural equation modeling,” in *A Second Course in Structural Equation Modeling*, eds G. R. Hancock and R. O. Mueller (Charlotte, NC: Information Age), 439–492.

Fong, D. Y., Ho, S. Y., and Lam, T. H. (2010). Evaluation of internal reliability in the presence of inconsistent responses. *Health Qual. Life Outcomes* 8:27. doi: 10.1186/1477-7525-8-27

Fonsêca, P. N., Couto, R. N., Melo, C. C., Amorim, L. A. G., and Pessoa, V. S. A. (2018b). Uso de redes sociais e solidão: evidências psicométricas de escalas [Use of social networks and loneliness: psychometric evidence of scales]. *Arq. Bras. Psicol.* 70, 198–212.

Fonsêca, P., Couto, R., Melo, C., Machado, M., and Souza Filho, J. F. (2018a). Escala de uso problemático de internet en estudiantes universitarios: evidencias de validez y fiabilidad [Scale of problematic internet use in university students: evidence of validity and reliability]. *Cienc. Psicol.* 12:223. doi: 10.22235/cp.v12i2.1686

García-Umaña, A., and Córdoba, E. (2020). Validación de escala MPPUS-A sobre el uso problemático del smartphone [Validation of Scale MPPUS-A on the problematic use of the smartphone]. *Pixel-Bit* 57, 173–189. doi: 10.12795/pixelbit.2020.i57.07

Gil, F., Oberst, U., Del Valle, G., and Chamarro, A. (2015). Nuevas tecnologías - ¿Nuevas patologías? El Smartphone y el fear of missing out [New technologies - New pathologies? The smartphone and the fear of missing out]. *Aloma* 33, 77–83. doi: 10.51698/aloma.2015.33.2.77-83

Godinho, A., Kushnir, V., and Cunningham, J. A. (2016). Unfaithful findings: identifying careless responding in addictions research. *Addiction* 111, 955–956. doi: 10.1111/add.13221

Green, S. B., and Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: an alternative to coefficient alpha. *Psychometrika* 74, 155–167. doi: 10.1007/s11336-008-9099-3

Griffiths, M. D. (1996). Behavioural addictions: An issue for everybody? *J. Work. Learn.* 8, 19–25.

Griffiths, M. D. (2005a). A ‘components’ model of addiction within a biopsychosocial framework. *J. Subst. Use* 10, 191–197. doi: 10.1080/14659890500114359

Griffiths, M. D. (2005b). Adiccción a los videojuegos: una revisión de la literatura. *Psicol. Conductual* 13, 445–462.

Guertler, D., Broda, A., Bischof, A., Kastirke, N., Meerkerk, G. J., John, U., et al. (2014). Factor structure of the compulsive internet use scale. *Cyberpsychol. Behav. Soc. Netw.* 17, 46–51. doi: 10.1089/cyber.2013.0076

Guttman, L. (1944). A basis for scaling qualitative data. *Am. Sociol. Rev.* 9, 139–150. doi: 10.2307/2086306

Hagquist, C. (2019). Explaining differential item functioning focusing on the crucial role of external information – an example from the measurement of adolescent mental health. *BMC Med. Res. Methodol.* 19:185. doi: 10.1186/s12874-019-0828-3

Herrero, J., and Gracia, E. (2007). Una medida breve de la sintomatología depresiva (CESD-7) [A brief measure of depressive symptomatology CESD-7]. *Salud Ment.* 30, 40–46.

Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., and Donohew, R. L. (2002). Reliability and validity of a brief measure of sensation seeking. *Pers. Individ. Differ.* 32, 401–414. doi: 10.1016/S0191-8869(01)00032-0

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Huang, J. L., Liu, M., and Bowling, N. A. (2015). Insufficient effort responding: examining an insidious confound in survey data. *J. Appl. Psychol.* 100, 828–845. doi: 10.1037/a0038510

Hussain, Z., and Griffiths, M. D. (2018). Problematic social networking site use and comorbid psychiatric disorders: a systematic review of recent large-scale studies. *Front. Psychiatry* 9:686. doi: 10.3389/fpsyt.2018.00686

Iovu, M. B., Runcan, R., Runcan, P. L., and Andrioni, F. (2020). Association between Facebook use, depression and family satisfaction: a cross-sectional study of romanian youth. *Iran J. Public Health* 49, 2111–2119. doi: 10.18502/ijph.v49i11.4728

Jia, Y., Konold, T. R., Cornell, D., and Huang, F. (2018). The impact of validity screening on associations between self-reports of bullying victimization and student outcomes. *Educ. Psychol. Meas.* 78, 80–102. doi: 10.1177/0013164416671767

Joo, H., Aguinis, H., and Bradley, K. J. (2017). Not all nonnormal distributions are created equal: improved theoretical and measurement precision. *J. Appl. Psychol.* 102, 1022–1053. doi: 10.1037/apl0000214

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., and Rosseel, Y. (2021). *SemTools: useful tools for structural equation modeling. R package version 0.5-5*. Vienna: R Core Team.

Karim, A. K. R., and Nigar, N. (2014). The internet addiction test: assessing its psychometric properties in Bangladeshi culture. *Asian J. Psychiatr.* 10, 75–83. doi: 10.1016/j.ajp.2013.10.011

Kelley, K. (2019). *MBESS: the MBESS R package. R package version 4.6.0*. Vienna: R Core Team.

Kelley, K., and Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychol. Methods* 21, 69–92. doi: 10.1037/a0040086

King, K. M., Kim, D. S., and McCabe, C. J. (2018). Random responses inflate statistical estimates in heavily skewed addictions data. *Drug Alcohol Depend.* 183, 102–110. doi: 10.1016/j.drugalcdep.2017.10.033

Koedel, C., and Betts, J. (2010). Value added to what? how a ceiling in the testing instrument influences value-added estimation. *Educ. Finance Policy* 5, 54–81. doi: 10.2139/ssrn.1261014

Koller, I., Levenson, M. R., and Glück, J. (2017). What do you think you are measuring? A mixed-methods procedure for assessing the content validity of test items and theory-based scaling. *Front. Psychol.* 8:126. doi: 10.3389/fpsyg.2017.00126

Koo, H. J., and Kwon, J. H. (2014). Risk and protective factors of internet addiction: a meta-analysis of empirical studies in Korea. *Yonsei Med. J.* 55, 1691–1711. doi: 10.3349/ymj.2014.55.6.1691

Korkmaz, S., Goksuluk, D., and Zararsiz, G. (2014). MVN: an R package for assessing multivariate normality. *R J.* 6, 151–162. doi: 10.32614/RJ-2014-031

Laconi, S., Urbán, R., Kaliszewska-Czeremska, K., Kuss, D. J., Gnisci, A., Sergi, I., et al. (2019). Psychometric evaluation of the nine-item problematic internet use questionnaire (PIUQ-9) in nine European samples of internet users. *Front. Psychiatry* 10:136. doi: 10.3389/fpsyt.2019.00136

Lam-Figueroa, N., Contreras-Pulache, H., Mori-Quispe, E., Nizama- Valladolid, M., Gutiérrez, C., Hinostroza-Camposano, W., et al. (2011). Adicción a internet: desarrollo y validación de un instrumento en escolares adolescentes de Lima, Perú [Internet adiction: development and validation of an instrument in adolescent scholars in Lima, Perú]. *Rev. Peru. Med. Exp. Salud Pública* 28, 462–469. doi: 10.1590/s1726-46342011000300009

Linacre, J. M. (1999). Investigating rating scale category utility. *J. Outcome Meas.* 3, 103–122.

Linacre, J. M. (2002a). Optimizing rating scale category effectiveness. *J. Appl. Meas.* 3, 85–106.

Linacre, J. M. (2002b). What do infit and outfit, mean-square and standardized mean? *Rasch Meas. Trans.* 16:878.

Liu, Y., Wu, A. D., and Zumbo, B. D. (2010). The impact of outliers on cronbach's coefficient alpha estimate of reliability: ordinal/rating scale item responses. *Educ. Psychol. Meas.* 70, 5–21. doi: 10.1177/0013164409344548

Lorca, M. M., and Sanz, C. A. (2003). Búsqueda de sensaciones y autoconcepto, asertividad y consumo de drogas existe relación? [Sensation seeking, self-concept, assertiveness and drug use. is there a relation?]. *Adicciones* 15, 145–158. doi: 10.20882/adicciones.438

Loss, A., Guerra, V. M., and Souza, M. L. (2020). Associação entre uso de Internet, autoconsciência ruminativa e diferenças de gênero em universitarios [Association between internet use, ruminative self-consciousness, and gender differences in university students]. *Av. Psicol. Latinoam.* 39, 1–14.

Lupano, M. L., and Castro, A. (2021). Rasgos de personalidad, bienestar y malestar psicológico en usuarios de redes sociales que presentan conductas disruptivas online [Personality traits, well-being and psychological distress in users who present online disruptive behavior]. *Rev. Psicol. Cienc. Afines* 38, 7–23. doi: 10.16888/interd.2021.38.2.1

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proc. Natl. Inst. Sci. India* 12, 49–55.

Mair, P., Hatzinger, R., and Maier, M. J. (2021). *eRm: extended rasch modeling. R package version 1.0-2*. Vienna: R Core Team.

Mangiafico, S. (2021). *rcompanion: functions to support extension education program evaluation. R package version 2.4.0*. Vienna: R Core Team.

Marco, C., and Chóliz, M. (2013). Tratamiento cognitivo-conductual en un caso de adicción a internet y videojuegos [Cognitive-behavioral treatment in case of internet and video game addiction]. *Int. J. Psychol. Ther.* 13, 125–141.

Marcos, M. (2020). Juego Online: tratamiento de un caso de adicción a apuestas deportivas [Online gambling: a treatment for sports betting addiction]. *Liberabit* 26:e399. doi: 10.24265/liberabit.2020.v26n2.04

Matute, H. (2016). "Adicción, abuso o uso problemático de Internet? [Addiction, abuse or problematic use of the Internet?]," in *Abuso de Internet: Antesala Para la Adicción al Juego de Azar Online*, ed. E. En Echeburúa (Madrid: Pirámide), 19–34.

Maydeu-Olivares, A. (2017). Maximum likelihood estimation of structural equation models for continuous data: standard errors and goodness of fit. *Struct. Equ. Model.* 24, 383–394. doi: 10.1080/10705511.2016.1269606

Maydeu-Olivares, A., and Coffman, D. L. (2006). Random intercept item factor analysis. *Psychol. Methods* 11, 344–362. doi: 10.1037/1082-989x.11.4.344

Maydeu-Olivares, A., and Steenkamp, J.-B. E. M. (2018). *An integrated procedure to control for common method variance in survey data using random intercept factor analysis models*. Available online at: https://www.academia.edu/36641946/An_integrated_procedure_to_control_for_common_method_variance_in_survey_data_using_random_intercept_factor_analysis_models. (accessed December 22, 2020)

Maydeu-Olivares, A., Shi, D., and Rosseel, Y. (2018). Assessing fit in structural equation models: a monte-carlo evaluation of RMSEA versus SRMR confidence intervals and tests of close fit. *Struct. Equ. Model.* 25, 389–402. doi: 10.1080/10705511.2017.1389611

Meade, A., and Craig, B. (2012). Identifying careless responses in survey data. *Psychol. Methods* 17, 437–455. doi: 10.1037/a0028085

Medrano, J. L. J., Rosales, F. L., and Loving, R. D. (2017). Conducta adictiva a las redes sociales y su relación con el uso problemático del móvil [Addictive behavior to social network sites and its relationship with the problematic use of the mobile phone]. *Acta Investig. Psicol.* 7, 2832–2838. doi: 10.1016/j.aipprr.2017.11.001

Meerkerk, G. J., Van Den Eijnden, R. J., Vermulst, A. A., and Garretsen, H. F. (2009). The Compulsive Internet Use Scale (CIUS): some psychometric properties. *Cyberpsychol. Behav.* 12, 1–6. doi: 10.1089/cpb.2008.0181

Merino-Soto, C., and Angulo-Ramos, M. (2020). Validity induction: Comments on the study of Compliance Questionnaire for Rheumatology. *Rev. Col. Reum.* 28, 312–313. doi: 10.1016/j.rcreu.2020.05.005

Merino-Soto, C., and Calderón-De la Cruz, G. (2018). Validez de estudios peruanos sobre estrés y burnout [Validity of Peruvian studies on stress and burnout]. *Rev Peruana Med. Exp.* 35, 353–354. doi: 10.17843/rpmesp.2018.352.3521

Merino-Soto, C., and Salas Blas, E. (2018). Brief sensation seeking scale: latent structure of 8-item and 4-item versions in peruvian adolescents. *Adicciones* 30, 41–53. doi: 10.20882/adicciones.842

Meyer, J. F., Faust, K. A., Faust, D., Baker, A. M., and Cook, N. E. (2013). Careless and random responding on clinical and research measures in the addictions: a concerning problem and investigation of their detection. *Int. J. Ment. Health Addict.* 11, 292–306. doi: 10.1007/s11469-012-9410-5

Mizumoto, A. (2015). *Langtest (Version 1.0)*. Available online at: http://langtest.jp. (accessed March 1, 2021)

Müller, K. W., Dreier, M., Beutel, M. E., Duven, E., Giralt, S., and Wölfling, K. (2016). A hidden type of internet addiction? Intense and addictive use of social networking sites in adolescents. *Comput. Hum. Behav.* 55, 172–177. doi: 10.1016/j.chb.2015.09.007

Nadal, R. (2008). La búsqueda de sensaciones y su relación con la vulnerabilidad a la adicción y al estrés [Novelty-seeking: its relationship with vulnerability to addiction and stress]. *Adicciones* 20, 59–72. doi: 10.20882/adicciones.289

Olsen, J. (1998). IEA European Questionnaire Group. Epidemiology deserves better questionnaires. *Int. J. Epidemiol.* 27:935.

Ord, J. K. (2006). "Pearson system of distributions," in *Encyclopedia of Statistical Sciences*, eds S. Kotz, N. Balakrishnan, C. B. Read, B. Vidakovic, and N. L. Johnson (New York, NY: John Wiley & Sons), 6036–6040.

Peris, M., Maganto, C., and Garaigordobil, M. (2018). Escala de riesgo de adicción-adolescente a las redes sociales e internet: fiabilidad y validez (ERA-RSI) [Scale of risk of addiction-adolescent to social networks and internet: reliability and validity]. *Rev. Psicol. Clín. Con Niños Adolesc.* 5, 30–36. doi: 10.21134/rpcna.2018.05.2.4

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* 88, 879–903. doi: 10.1037/0021-9010.88.5.879

Puerta-Cortés, D., and Carbonell, X. (2013). Problematic Internet use in a sample of Colombian university students. *Av. Psicol. Latinoam.* 31, 620–631.

Rodriguez, A. P., and Fernandez, A. (2014). Relación entre el tiempo de uso de las redes sociales en internet y la salud mental en adolescentes colombianos [Relationship between the time spent on internet social networking and mental health in Colombian adolescents]. *Acta Colomb. Psicol.* 17, 131–140. doi: 10.14718/ACP.2014.17.1.13

Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

Ruiz-Ruano, A. M., López-Salmerón, M. D., and López Puga, J. (2020). Experiential avoidance and excessive smartphone use: a Bayesian approach. *Adicciones* 32, 116–127. doi: 10.20882/adicciones.1151

Sahin, C. (2018). Social media addiction scale-student form: the reliability and validity study. *Turk. Online J. Educ. Technol.* 17, 169–182.

Salas, E. (2014). Adicciones psicológicas y los nuevos problemas de salud [Psychological addictions and new health problems]. *Cultura* 28, 111–146.

Salas, E., and Escurra, M. (2014). Uso de redes sociales entre universitarios limeños [Use of social networks among Lima university students]. *Rev. Peru. Psicol. Trab. Soc.* 3, 75–90.

Salas-Blas, E. (2019a). "Las adicciones comportamentales [Behavioral addictions]," in *Drogas: Sujeto, Sociedad y Cultura*, ed. C. Rojas-Jara (Talca: Nueva Mirada Ediciones), 123–138.

Salas-Blas, E. (2019b). *Patrones de uso y abuso de las TIC entre adolescentes de Lima y Arequipa. Percepción de los riesgos [ICT use and abuse patterns among adolescents in Lima and Arequipa. Perception of risks]. Informe final. Fundación MAPFRE.* Available online at: https://app.mapfre.com/documentacion/publico/es/catalogo_imagenes/grupo.do?path=1102847. (accessed January 15, 2021).

Salas-Blas, E., Copez-Lonzoy, A., and Merino-Soto, C. (2020). ¿Realmente es demasiado corto? Versión breve del cuestionario de adicción a redes sociales (ARS-6) [Is It Really Too Short? Brief Version of the Social Network Addiction Questionnaire (ARS-6). *Health Addict.* 20, 105–118. doi: 10.21134/haaj.v20i2.536

Salazar-Concha, C., Barros, D., and Quinn, J. (2021). Comportamiento en el uso de redes sociales en estudiantes de enseñanza media: los casos de un colegio particular y uno público en Chile [Behavior in the use of social networks in high school students: the cases of a private and a public school in Chile]. *Rev. Ibér. Sist. Tecnol. Inf. E* 42, 506–519.

Santamaria, A. C., and Vallejos-Flores, M. (2021). Diseño y validez de la Escala de Adicción a Instagram de Bergen (BIAS) en adultos peruanos [Design and validity of the Bergen Instagram Addiction Scale (BIAS) in Peruvian adults]. *Propós. Represent.* 9:e973. doi: 10.20511/pyr2021.v9n1.973

Santor, D. A., and Coyne, J. C. (1997). Shortening the CES–D to improve its ability to detect cases of depression. *Psychol. Assess.* 9, 233–243. doi: 10.1037/1040-3590.9.3.233

Saris, W. E., Satorra, A., and van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Struct. Equ. Model.* 16, 561–582. doi: 10.1080/10705510903203433

Šimkovic, M., and Träuble, B. (2019). Robustness of statistical methods when measure is affected by ceiling and/or floor effect. *PLoS One* 14:e0220889. doi: 10.1371/journal.pone.0220889

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., and Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Med. Res. Methodol.* 8:33. doi: 10.1186/1471-2288-8-33

Sociedad Mexicana de Psicología (2010). *Código Ético del Psicólogo [Ethical Code of the Psychologist]*. México: Editorial Trillas.

Stark, S., Chernyshenko, O. S., and Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy. *J. Appl. Psychol.* 91, 1292–1306. doi: 10.1037/0021-9010.91.6.1292

Steenkamp, J.-B. E. M., and Maydeu-Olivares, A. (2021). An updated paradigm for evaluating measurement invariance incorporating common method variance and its assessment. *J. Acad. Mark. Sci.* 49, 5–29. doi: 10.1007/s11747-020-00745-z

Stephenson, M. T., Velez, L. F., Chalela, P., Ramirez, A., and Hoyle, R. H. (2007). The reliability and validity of the Brief Sensation Seeking Scale (BSSS-8) with young adult Latino workers: implications for tobacco and alcohol disparity research. *Addiction* 102, 79–91. doi: 10.1111/j.1360-0443.2007.01958.x

Tao, R., Huang, X., Wang, J., Zhang, H., Zhang, Y., and Li, M. (2010). Proposed diagnostic criteria for internet addiction. *Addiction* 105, 556–564. doi: 10.1111/j.1360-0443.2009.02828.x

Taras, V., and Kline, T. (2010). Scale validation via quantifying item validity using the Dm Index. *Psychol. Rep.* 107, 535–546. doi: 10.2466/03.PR0.107.5.535-546

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *J. Clin. Epidemiol.* 60, 34–42. doi: 10.1016/j.jclinepi.2006.03.012

Tomczak, M., and Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends Sports Sci.* 1, 1–25.

Trafimow, D., Wang, T., and Wang, C. (2019). From a sampling precision perspective, skewness is a friend and not an enemy! Educ. *Psychol. Meas.* 79, 129–150. doi: 10.1177/0013164418764801

Turel, O., and Serenko, A. (2012). The benefits and dangers of enjoyment with social networking websites. *Eur. J. Inf. Syst.* 21, 512–528. doi: 10.1057/ejis.2012.1

Valencia-Ortiz, R., Cabero-Almenara, J., Ruiz, U. G., and Robles, B. F. (2021). Problemática de estudio e investigación de la adicción a las redes sociales online en jóvenes y adolescentes [Study and research problems of addiction to online social networks in young people and adolescents]. *Rev. Tecnol. Cienc. Educ.* 2021, 99–125. doi: 10.51302/tce.2021.573

Vallejos-Flores, M. A., Copez-Lonzoy, A., and Capa-Luque, W. (2018). ¿Hay alguien en línea: Validez y confiabilidad de la versión en español de la Bergen Facebook Addiction Scale (BFAS) en universitarios [Is there anyone online? Validity and reliability of the spanish version of the Bergen Facebook addiction scale (BFS) in university students]. *Health Addict. Salud Drog.* 18, 175–184. doi: 10.21134/haaj.v18i2.394

van der Ark, A. (2012). New developments in mokken scale analysis in R. *J. Stat. Softw.* 48, 1–27. doi: 10.18637/jss.v048.i05

Vilca, L. W., and Vallejos, M. (2015). Construction of the risk of addiction to social networks scale (Cr.A.R.S.). *Comput. Hum. Behav.* 48, 190–198. doi: 10.1016/j.chb.2015.01.049

Wilcox, R. R. (2012). *Introduction to Robust Estimation and Hypothesis Testing*. San Diego, CA: Elsevier.

World Medical Association (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama* 310, 2191–2194. doi: 10.1001/jama.2013.281053

Wright, B. D., and Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Meas. Trans.* 8:370.

Yentes, R. D., and Wilhelm, F. (2021). *Careless: procedures for computing indices of careless responding. R package version 1.2.1*. Vienna: R Core Team.

Young, K. S. (1998). Internet addiction: the emergence of a new clinical disorder. *CyberPsychol. Behav.* 1, 237–244. doi: 10.1089/cpb.1998.1.237

Zhang, Y., Mei, S., Li, L., Chai, J., Li, J., and Du, H. (2015). The relationship between impulsivity and internet addiction in Chinese college students: a moderated mediation analysis of meaning in life and self-esteem. *PLoS One* 10:e0131597. doi: 10.1371/journal.pone.0131597

Zhu, W. (2002). A confirmatory study of Rasch-based optimal categorization of a rating scale. *J. Appl. Meas.* 3, 1–15.

Zhu, W., Updyke, W. F., and Lewandowski, C. (1997). Post-hoc Rasch analysis of optimal categorization of an ordered-response scale. *J. Outcome Meas.* 1, 286–304.

Zijlstra, W. P., van der Ark, L. A., and Sijtsma, K. (2007). Outlier detection in test and questionnaire data. *Multivar. Behav. Res.* 42, 531–555. doi: 10.1080/00273170701384340

Zijlstra, W. P., van der Ark, L. A., and Sijtsma, K. (2011). Outliers in questionnaire data: can they be detected and should they be removed? *J. Educ. Behav. Stat.* 36, 186–212. doi: 10.3102/1076998610366263

Zijlstra, W. P., Van der Ark, L. A., and Sijtsma, K. (2013). Discordancy tests for outlier detection in multi-item questionnaires. *Methodology* 9, 69–77. doi: 10.1027/1614-2241/a000056

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 13 January 2022
doi: 10.3389/fpsyg.2021.784222

The Infertility-Related Stress Scale: Validation of a Brazilian–Portuguese Version and Measurement Invariance Across Brazil and Italy

Giulia Casu[1], *Victor Zaia*[2,3]*, *Erik Montagna*[2], *Antonio de Padua Serafim*[4,5], *Bianca Bianco*[2,3], *Caio Parente Barbosa*[2,3] and *Paola Gremigni*[1]

[1] Department of Psychology, University of Bologna, Bologna, Italy, [2] Postgraduate Program in Health Sciences, Centro Universitário FMABC, Santo André, Brazil, [3] Ideia Fértil Institute of Reproductive Health, Centro Universitário FMABC, Santo André, Brazil, [4] Neuropsychology Unit, Department and Institute of Psychiatry, University of São Paulo School of Medicine, São Paulo, Brazil, [5] Department of Psychology, Methodist University of São Paulo, São Bernardo do Campo, Brazil

OPEN ACCESS

Edited by:
Cesar Merino-Soto,
Universidad de San Martin de Porres, Peru

Reviewed by:
Leire Aperribai,
University of the Basque Country, Spain
Lucia Bonassi,
Bolognini Hospital, Italy

***Correspondence:**
Victor Zaia
victor.zaia@fmabc.br;
victorzaia@gmail.com

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *27 September 2021*
Accepted: *24 December 2021*
Published: *13 January 2022*

Citation:
Casu G, Zaia V, Montagna E, de Padua Serafim A, Bianco B, Barbosa CP and Gremigni P (2022) The Infertility-Related Stress Scale: Validation of a Brazilian–Portuguese Version and Measurement Invariance Across Brazil and Italy. Front. Psychol. 12:784222. doi: 10.3389/fpsyg.2021.784222

Infertility constitutes an essential source of stress in the individual and couple's life. The Infertility-Related Stress Scale (IRSS) is of clinical interest for exploring infertility-related stress affecting the intrapersonal and interpersonal domains of infertile individuals' lives. In the present study, the IRSS was translated into Brazilian–Portuguese, and its factor structure, reliability, and relations to sociodemographic and infertility-related characteristics and depression were examined. A sample of 553 Brazilian infertile individuals (54.2% female, mean aged 36 $\pm$ 6 years) completed the Brazilian–Portuguese IRSS (IRSS-BP), and a subsample of 222 participants also completed the BDI-II. A sample of 526 Italian infertile individuals (54.2% female, mean aged 38 $\pm$ 6 years) was used to test for the IRSS measurement invariance across Brazil and Italy. Results of exploratory structural equation modeling (ESEM) indicated that a bifactor solution best represented the structure underlying the IRSS-BP. Both the general and the two specific intrapersonal and interpersonal IRSS-BP factors showed satisfactory levels of composite reliability. The bifactor ESEM solution replicated well across countries. As evidence of relations to other variables, female gender, a longer duration of infertility, and higher depression were associated with higher scores in global and domain-specific infertility-related stress. The findings offer initial evidence of validity and reliability of the IRSS-BP, which could be used by fertility clinic staff to rapidly identify patients who need support to deal with the stressful impact of infertility in the intrapersonal and interpersonal life domains, as recommended by international guidelines for routine psychosocial care in infertility settings.

Keywords: infertility-related stress, validation, exploratory structural equation modeling, bifactor model, measurement invariance, depressive symptoms

INTRODUCTION

According to the World Health Organization (World Health Organization [WHO], 2006), approximately 10% of couples of reproductive age worldwide have difficulties achieving pregnancy. Infertility has been defined as the absence of conception after 12 months of regular unprotected sexual intercourse (Zegers-Hochschild et al., 2017). It is estimated that more than 48 million couples worldwide suffer from infertility (Mascarenhas et al., 2012), and most of them live in developing countries (Ombelet et al., 2008). In Brazil, it is estimated that 15–20% of couples of reproductive age have some infertility problems (Instituto Brasileiro de Geografia e Estatística [IBGE], 2010).

Infertile individuals and couples experience considerable stress because of failing to achieve a meaningful life goal such as parenthood and its accompanying social stigma (Loke et al., 2012; Öztürk et al., 2021a). Stress can affect immune system activity and thus lead to physical and mental vulnerability (Segerstrom and Miller, 2004). Implications of infertility include physical symptoms, reduced psychological well-being and quality of life, feelings of guilt and shame, use of negative avoidance coping strategies, and social isolation (Rockliff et al., 2014; Luk and Loke, 2015; Rooney and Domar, 2016; Swanson and Braverman, 2021). Recently, a prevalence of 21–52% has been reported for depressive symptoms among infertile women, which is well above rates in the general population (Kiani et al., 2021). Women are the most affected by the infertile experience, consistently showing greater social vulnerability, stress and emotional distress, and lower quality of life than men (Patel et al., 2018; Casu et al., 2019; Ha and Ban, 2020).

To solve their infertility problems, more than half of infertile couples seek medical care like assisted reproductive technology (ART) treatments (Boivin et al., 2007). ART treatments constitute an additional source of stress for infertile couples due to expensive, time-consuming, invasive, and physically demanding procedures, along with the uncertainty of the outcomes (Öztürk et al., 2021b). Indeed, infertile patients, and especially women, report that ART treatments are physically and emotionally exhausting (Arya and Dibb, 2016; Anaman-Torgbor et al., 2021; Öztürk et al., 2021b), which is a major cause of premature treatment discontinuation by 30% of couples (Pedro et al., 2017; Domar et al., 2018). Altogether, the specific stress associated with infertility and its treatment can have long-lasting psychosocial consequences on infertile individuals and couples (Schmidt, 2009).

The stress specific to infertility, namely infertility-related stress, has been conceptualized as the burden that the infertile condition imposes on different life domains and entails areas of patient concern such as social, relationship, and sexual concerns, need for parenthood, and negative evaluation of childlessness (Newton et al., 1999; Schmidt et al., 2005). Thus, infertility-related stress has effects at the individual intrapersonal and interpersonal levels. Screening infertile patients for their levels of infertility-related stress can be especially useful in ART settings. Indeed, identifying patient needs by fertility staff has been recommended by the European Society of Human Reproduction and Embryology (ESHRE) guidelines for routine psychosocial care to reduce stress and improve patient well-being and compliance with treatment (Gameiro et al., 2015).

Various measures have been developed to assess infertility-related stress, such as the 46-item Fertility Problem Inventory (FPI; Newton et al., 1999) and the 36-item Fertility Quality of Life Tool (FertiQoL; Boivin et al., 2011). However, time constraints in health settings, such as fertility clinics, make the use of brief measures advisable to minimize patient and staff burden and provide a time-efficient assessment (Ziegler et al., 2014). A brief self-report questionnaire has been recently developed to measure intrapersonal and interpersonal infertility-related stress in infertile Italian individuals, namely the Infertility-Related Stress Scale (IRSS) (Casu and Gremigni, 2016). The IRSS intrapersonal dimension refers to one's identity and resources for mind and body well-being as affected by infertility stress and includes aspects such as mental and physical health, intimacy, leisure, and life satisfaction. The interpersonal dimension refers to one's social roles, rights, and responsibilities and includes aspects like relationships with others in the social and familial environment and work performance. In the original validation study on a sample of 597 infertile women and men turning to ART (Casu and Gremigni, 2016), the IRSS showed evidence of good psychometric properties. Confirmatory factor analysis (CFA) supported the proposed two-correlated factor model of infertility-related stress impacting intrapersonal and interpersonal life domains. The two latent variables were strongly correlated (r = 0.72). The intrapersonal and interpersonal dimensions showed adequate internal consistency (Cronbach's αs of 0.89 and 0.87, respectively) and test-retest reliability over 4 weeks (intraclass correlation coefficients of 0.89 and 0.86, respectively). As for evidence of relations to other variables, in both women and men, the intrapersonal dimension was strongly (rs between 0.43 and 0.55), and the interpersonal dimension moderately (rs between 0.29 and 0.40) correlated with measures of anxiety and depression. Also, infertility-related stress was higher in the intrapersonal than in the interpersonal area of life in both genders, and women scored strongly higher than men in the intrapersonal domain (Casu and Gremigni, 2016).

Therefore, the IRSS appears to be a promising tool that might assist fertility staff in screening for infertility-related stress and identifying the areas of life most disrupted by the infertile condition. The IRSS has been used previously in the Brazilian research context, showing high levels of internal consistency for its global score, which correlated negatively with spirituality, quality of life, and perceived social support scores, and positively with avoidance coping scores (Casu et al., 2018, 2019). However, a Brazilian–Portuguese version has not been rigorously validated yet, and it is unclear whether its factor structure reflects a cross-cultural pattern (Casu and Gremigni, 2016).

This study aimed to validate a Brazilian–Portuguese translation of the IRSS (IRSS-BP) by testing for its factor structure, reliability, and relations to sociodemographic and infertility-related characteristics and depression as evidence of concurrent validity. To investigate the IRSS-BP factor structure, exploratory structural equation modeling (ESEM; Asparouhov and Muthén, 2009) was preferred over CFA. CFA unrealistically assumes that each item represents its designated construct exclusively, forcing items to load on only one factor and constraining all cross-loadings to zero (Marsh et al., 2014). To overcome the limitations of CFA, ESEM has been recommended, which combines the flexibility of exploratory factor analysis with the advantages of CFA (e.g., estimation of goodness-of-fit indices and the possibility of multigroup analysis) while offering a more realistic representation of the data (Morin et al., 2013; Marsh et al., 2014). Using ESEM, we tested both the two-correlated factor model proposed in the IRSS original validation study and a bifactor model with one general factor and two specific factors. Indeed, the strong correlation between the intrapersonal and interpersonal domains of infertility-related stress found in the original validation study may suggest the presence of a global factor underlying all IRSS indicators (Morin et al., 2016). Measurement invariance of the IRSS across Brazilian and Italian samples was also tested. Evidence of measurement invariance would indicate that Brazilian infertile individuals conceptualize and evaluate infertility-related stress similarly to Italian ones (Steenkamp and Baumgartner, 1998). Concerning sociodemographic characteristics considered in concurrent validity testing, previous studies reported that risk factors for higher infertility-related stress or lower psychological health included female gender (Ying et al., 2015), older age (Lakatos et al., 2017), and low educational level (Zaidouni et al., 2018). About infertility-related characteristics, suffering from primary infertility (i.e., no prior pregnancies), a longer duration of the infertility problem, and a diagnosis of female factor infertility were associated with higher infertility-related stress (Patel et al., 2016; Zaidouni et al., 2018). Concerning depression, infertility-related stress was found to significantly predict depressive symptoms among infertile women and men (Zurlo et al., 2020).

MATERIALS AND METHODS

Participants and Procedure

After approval by the institutional review board at both Brazilian and Italian institutions, participants at both sites were invited to participate and explained the scope of the study. Participation was voluntary and anonymous. All participants provided informed consent, and the study was conducted following the Declaration of Helsinki.

Inclusion criteria for both Brazilian and Italian participants were being 18 years or older, able to fill out a questionnaire in Brazilian-Portuguese/Italian language without help, having a diagnosis of infertility defined as the failure to achieve a clinical pregnancy after 12 months or more of regular unprotected sexual intercourse (Zegers-Hochschild et al., 2017), and being currently in or seeking ART treatment.

Participants who met the inclusion criteria in Brazil were invited to complete a paper-and-pencil questionnaire which included the Brazilian–Portuguese version of the IRSS and items on sociodemographic (i.e., gender, age, education) and infertility-related characteristics. Infertility-related characteristics included duration of infertility (coded as 1–2, 3–4, 5–6, and >6 years), infertility type (i.e., primary infertility, defined as having never conceived despite at least 12 months of attempting conception, or secondary infertility, defined as having had at least one prior conception but being subsequently unable to conceive after at least 12 months of attempting conception), and infertility diagnosis (coded as no diagnosis, male factor, female factor, both male and female factor, and unexplained). A randomized subsample of participants also completed the Brazilian–Portuguese version of the Beck Depression Inventory-II (BDI-II; Beck et al., 1996; Gomes-Oliveira et al., 2012). Participants who met the inclusion criteria in Italy were invited to fill in the original Italian IRSS and asked questions on age, education, infertility type, and infertility diagnosis.

Brazilian participants were recruited at a fertility clinic in the São Paulo metropolitan region, Brazil. Italian participants were recruited at two private fertility clinics in the metropolitan area of Bologna, Italy.

Measures

Infertility-Related Stress Scale

The IRSS (Casu and Gremigni, 2016) is a 12-item self-report measure to assess the amount of stress the infertility problem places on different aspects of life. It consists of two 6-item subscales referring to the intrapersonal (e.g., mental well-being) and interpersonal (e.g., friends) domains of life. Each item is rated on a 7-point scale from 1 (not at all) to 7 (a great deal). McDonald's ω of the Italian IRSS in the present study (n = 526) was 0.89 for both the intrapersonal (95% CI 0.88–0.91) and interpersonal (95% CI 0.87–0.91) domains, and 0.93 (95%CI 0.92–0.94) for the total IRSS.

Two independent bilingual translators translated the Italian IRSS into Brazilian-Portuguese and then back-translated it into Italian. Discussion between the translators and the Italian-speaking researchers resolved any discrepancies between the original and back-translated versions. Semantic validation of the final Brazilian–Portuguese translation was then performed in two focus groups (Mayring, 2014) of infertile individuals (n = 6, 50% women, by group). All participants in the focus groups were undergoing infertility treatment and had different educational levels. The lowest educational level was 8 years of schooling (corresponding to compulsory education in Brazil), and the highest level was 21 years of schooling (corresponding to Ph.D.). Participants were asked for their opinion about the readability and clarity of the instrument. Only slight changes were proposed to enhance clarity of instructions, which were implemented by the second author who facilitated the focus groups. The final Brazilian-Portuguese version of the IRSS was reviewed and approved by all focus group participants and named IRSS-BP. The final IRSS-BP version did not show substantial differences from the initial one.

Beck Depression Inventory-II

The BDI-II (Beck et al., 1996) is a widely used 21-item self-report measure to assess cognitive, motivational, affective, and somatic symptoms of depression. For each item, respondents are asked to choose the statement that best describes their feelings in the past 2 weeks. Each item is scored 0–3, with higher scores indicating greater depression severity. We used the Brazilian–Portuguese validated version of the BDI-II (Gomes-Oliveira et al., 2012).

A total score of >10 was considered to differentiate between participants with below- and above-threshold levels of depression (Gomes-Oliveira et al., 2012). McDonald's ω in the present study ($n = 222$) was 0.89 (95% CI 0.87–0.91).

Data Analysis

Prior to psychometric analyses, outliers and careless responders were identified and removed from the dataset to improve data quality (Curran, 2016). Multivariate outliers were defined as any case with a Mahalanobis distance (D^2) above the critical χ^2 value of 32.91 (12 degrees of freedom, $p < 0.001$). Careless responders were detected using the inter-item standard deviation (ISD) to measure an individual's inconsistent responding. Participants with ISD values two standard deviations above the mean were considered careless responders (Marjanovic et al., 2015).

Preliminary analyses on the final dataset included item descriptive statistics, univariate and multivariate normality, and associations with sociodemographics (i.e., gender and age). At the univariate level, IRSS-BP items with skewness and kurtosis $< |2|$ were considered normally distributed. To test for multivariate normality, the Henze-Zirkler test was used. Associations of IRSS-BP items with gender and age were examined using point-biserial and product-moment correlations, respectively.

To investigate the factor structure of the IRSS-BP, ESEM with target rotation (Asparouhov and Muthén, 2009) was conducted. Target rotation enables ESEM to be used in a confirmatory way by allowing for an *a priori* specified configuration of indicators for each factor. In addition to the principal loadings, all cross-loadings are freely estimated in target rotation but targeted to be as close to zero as possible (Asparouhov and Muthén, 2009). Two ESEM models were tested: a first-order model with two correlated factors and a bifactor model. In the two-correlated factor model, each item loaded on intrapersonal and interpersonal factors. In the bifactor model, a general factor (G-Factor) and two specific factors (S-Factor intrapersonal and S-Factor interpersonal) were included, and each item loaded directly and simultaneously on the G-Factor and both the S-Factors. We used oblique target rotation in the first-order model and orthogonal target rotation in the bifactor model (Reise, 2012). In both models, loadings ≥ 0.30 were considered relevant. Model parameters were estimated using the robust maximum likelihood method, robust to violations of multivariate normality, and recommended for variables with five or more response categories (Rhemtulla et al., 2012). The following goodness-of-fit indices were used: χ^2, Satorra–Bentler scaled χ^2 statistic (S-B χ^2), root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR) ≤ 0.08, and comparative fit index (CFI) ≥ 0.95 (Hu and Bentler, 1999). In the case of non-optimal model fit, modification indices were examined to find the most parsimonious changes to the model to achieve an acceptable fit to the data. To identify the model to be retained, we considered parameter estimates (loadings and cross-loadings) in addition to model fit indices. According to Morin et al. (2016), the bifactor ESEM model should be preferred if the G-Factor and S-Factors are well-defined, and cross-loadings in the bifactor ESEM solution decrease compared to its first-order counterpart. McDonald's (1970) omega (ω) coefficient of composite reliability was also considered, with values above 0.70 and 0.50 being satisfactory for measures corresponding to first-order and bifactor models, respectively (Perreira et al., 2018). For both models, we also considered the proportion of item variance explained by the model: σ^2 error, σ^2 true related to the first-order factors in the first-order ESEM and the G-Factor and S-Factors in the bifactor ESEM, and σ^2 true related to cross-loadings (Perreira et al., 2018; Morin et al., 2020). For the bifactor model, the proportion of explained common variance (ECV) attributable to each factor was also computed (Reise et al., 2013).

Using data from the Brazilian and the Italian samples, a multigroup ESEM was conducted to test the measurement invariance of the most optimal model across the country. Increasingly restrictive models representing configural (invariant factor structure), metric/weak (invariant factor loadings), scalar/strong (invariant intercepts), strict (invariant uniqueness), latent variance-covariance matrix, and latent factor means invariance (Steenkamp and Baumgartner, 1998). Differences in fit between nested models were evaluated considering, in addition to the S-B χ^2 difference test (ΔS-B χ^2), changes in CFI (ΔCFI), RMSEA (ΔRMSEA), and SRMR (ΔSRMR). A $\Delta\text{CFI} \leq 0.010$ supplemented by a $\Delta\text{RMSEA} < 0.015$ or a $\Delta\text{SRMR} < 0.010$ were considered as indicative of a non-significant decrease in fit across models (Chen, 2007). If full measurement invariance did not hold, partial measurement invariance was considered by relaxing equality constraints on measurement parameters (Steenkamp and Baumgartner, 1998). The sample size of the Brazilian and Italian samples was established *a priori* as to have approximately 10 observations for each freely estimated parameter in the models (Kline, 2005).

Analysis of variance (ANOVA) was performed to examine cross-country differences in IRSS scores and to test for relations of the IRSS-BP with sociodemographics (i.e., gender and education), infertility-related characteristics (i.e., duration of infertility, infertility type, and infertility diagnosis), and depressive symptomatology levels. Repeated-measures ANOVA was used to investigate differences in scores across IRSS-BP dimensions. Pearson correlations were computed to test for the associations of age and BDI-II scores with IRSS-BP scores.

Interpretation of results was based on statistical significance ($p < 0.05$) and measures of effect size, with r of 0.10 considered small, 0.30 medium and 0.50 large, and d of 0.20 considered small, 0.50 medium and 0.80 large (Cohen, 1988). ESEM models were estimated using Mplus 8.4 (Muthén and Muthén, 1998–2017). All other analyses were performed with IBM SPSS 27 (SPSS Inc., Chicago, IL, United States).

RESULTS

Detection of Outliers and Careless Responders

In Brazil, of 700 invited patients, 570 (81.4%) met the inclusion criteria and completed the IRSS-BP. Fourteen cases (2.5%) had a D^2 above the critical χ^2 value and were flagged as multivariate outliers. Three cases (0.5%) had both a D^2 greater than the critical χ^2 and an ISD value two standard deviations ($SD = 0.66$) above the mean (M = 1.08) and were flagged as both outliers and careless responders. The Brazilian sample used in subsequent analyses thus comprised n = 553 infertile patients. In Italy, of 680 invited patients, 557 (81.9%) met the inclusion criteria and completed the IRSS. Eleven cases (2.0%) were flagged as multivariate outliers due to D^2 values greater than the critical χ^2, 12 cases (2.2%) had ISD values two standard deviations (SD = 0.60) above the mean (M = 1.25) and were considered careless responders, and 8 additional cases (1.4%) were flagged as both outliers and careless responders. Therefore, the Italian sample used in the analyses included n = 526 infertile patients.

Participants' Characteristics

Brazilian participants (n = 553) were 54.2% female; most participants were highly educated, having a degree or post-degree, and had primary infertility. About one third had unexplained infertility, while 14.1% had not completed standard infertility evaluations and thus had not a specific infertility diagnosis yet. About age and education, women (M = 35.24, SD = 5.24, range 18–54 years) were slightly younger than men (M = 37.35, SD = 6.52, range 23–63 years) [$F(1,548) = 17.72$, $p < 0.001$, d = 0.36] and a slightly larger proportion of women (80.3%) than men (72.3%) were highly educated [$\chi^2(1) = 4.48$, p = 0.03]. There were no gender differences in duration of infertility [$\chi^2(3) = 5.64$, p = 0.13], type of infertility [$\chi^2(1) = 1.38$, p = 0.24], or infertility diagnosis [$\chi^2(3) = 2.34$, p = 0.51]. The subsample of 222 participants who also completed the BDI-II was 60.4% female (n = 134), mean aged 34.56 years (SD = 5.47, range 23–54), and 88.7% highly educated. There were no differences in infertility-related characteristics between Brazilian participants who completed only the IRSS-BP and those who also responded the BDI-II.

Italian participants (n = 526) were 54.2% female, and women (M = 36.20, SD = 4.62, range 23–50 years) were strongly younger than men (M = 40.36, SD = 6.23, range 25–59 years) [$F(1,524) = 77.01$, $p < 0.001$, d = 0.77]. No other statistically significant gender differences were found in the Italian sample.

Comparisons between the Brazilian and Italian samples showed that Brazilian participants were slightly younger than the Italians (d = 0.32), a larger proportion of Brazilian than Italian participants were highly educated, had primary infertility and had unexplained infertility. In contrast, a smaller proportion was undiagnosed or had both male and female factors. Characteristics of participants are presented in **Table 1**.

TABLE 1 | Participants' characteristics.

Characteristic	Brazilians (n = 553)	Italians (n = 526)	Comparison
Gender, women, n (%)	300 (54.2)	285 (54.2)	$\chi^2(1) = 0.001$
Age, years, mean (SD, range)	36.21 (5.95, 18–63)	38.10 (5.80, 23–59)	$F(1,1074) = 28.01$*
Education, high, n (%)	424 (76.7)	232 (44.1)	$\chi^2(1) = 119.96$*
Infertility type, primary, n (%)	375 (67.8)	283 (53.8)	$\chi^2(1) = 22.24$*
Infertility diagnosis, n (%)		–	$\chi^2(4) = 82.45$*
Undiagnosed	78 (14.1)[a]	110 (20.9)[b]	
Male	140 (25.3)[a]	142 (27.0)[a]	
Female	119 (21.5)[a]	99 (18.8)[a]	
Both male and female	53 (9.6)[a]	118 (22.4)[b]	
Unexplained	163 (29.5)[a]	57 (10.8)[b]	
Infertility duration, n (%)			
1–2 years	188 (34.0)	–	
3–4 years	165 (29.8)	–	
5–6 years	121 (21.9)	–	
> 6 years	79 (14.3)	–	

Proportions with different superscript letters in the same row significantly differ at p < 0.05 (post hoc z-scores and Bonferroni correction).
**p < 0.001.*

Preliminary Analyses of Brazilian–Portuguese Infertility-Related Stress Scale Items

All IRSS-BP items followed a univariate normal distribution, with skewness and kurtosis < |2|. However, the Henze-Zirkler test indicated significant departures from multivariate normality (HZ = 11.39, p = 0.008). As for relations to demographic variables (i.e., sex and age), associations with gender were significant for almost all items but only weak (r_{pb} = –0.21 to –0.07). Pearson's correlations with age were predominantly non-significant and ranged between –0.11 and –0.01. Item descriptive statistics are shown in **Supplementary Table S1**.

Factor Structure

The goodness of fit of the two-correlated factor solution was below acceptable levels [$\chi^2(43) = 455.899$, S-B $\chi^2(43) = 332.264$, $p < 0.001$, RMSEA = 0.110 (90% CI 0.099–0.122), SRMR = 0.034, CFI = 0.917]. Inspection of modification indices suggested that allowing the uniqueness of two items (i.e., item 2 – Relatives with item 3 – In-laws) covary would improve model fit. This covariation makes substantive and theoretical sense, as blood relatives and in-laws have been reported as the main sources of social pressure by infertile individuals (Hasanpoor-Azghdy et al., 2015). Therefore, the model was respecified, including an argument for the correlated uniqueness. The goodness of fit of the respecified model improved significantly, ΔS-B $\chi^2(1) = 96.587$, $p < 0.001$, with goodness-of-fit indices indicating acceptable fit to the data [$\chi^2(42) = 264.564$, S-B $\chi^2(42) = 194.876$, RMSEA = 0.081 (90% CI 0.070–0.093), SRMR = 0.027, CFI = 0.956]. As shown in **Table 2**, the two-correlated factor, first-order ESEM solution resulted in well-defined factors. Loadings on the target factor were > 0.40 for both the intrapersonal ($M_{|\lambda|}$= 0.740) and intrapersonal ($M_{|\lambda|}$= 0.773) dimensions. Loadings on non-target factors were significant and > |0.20| for 4 out of the 12 possible cross-loadings, ranging from 0.201 to 0.346. Composite reliability estimates were adequate for both factors (ω = 0.88). A correlation of 0.72 between the two factors was observed. Such a strong correlation might indicate the existence of a more general factor tapping variation in responses across all IRSS-BP items, supporting the estimation of a bifactor model.

The goodness of fit of the bifactor ESEM solution was adequate, with all goodness-of-fit indices meeting the pre-established criteria [$\chi^2(33) = 201.028$, S-B $\chi^2(33) = 150.380$,

TABLE 2 | Standardized factor loadings and item uniquenesses for the first-order and bifactor ESEM solutions (n = 553).

	First-order ESEM			Bifactor ESEM			
	Intrapersonal(λ)	**Interpersonal(λ)**	**δ**	**G-Factor(λ)**	**Intrapersonal(λ)**	**Interpersonal(λ)**	**δ**
(1) Physical well-being	**0.714**	0.126	0.345	0.657	**0.466**	*0.031*	0.351
(4) Leisure and enjoyment	**0.445**	0.346	0.461	0.654	**0.302**	0.114	0.468
(5) Marital satisfaction	**0.617**	*0.106*	0.514	0.616	**0.341**	–0.132	0.486
(6) Mental well-being	**0.960**	–0.082	0.186	0.630	**0.656**	*0.013*	0.172
(9) Sexual pleasure	**0.752**	*0.037*	0.393	0.609	**0.477**	*–0.039*	0.400
(12) Global life satisfaction	**0.950**	–0.118	0.244	0.587	**0.650**	*0.004*	0.232
(2) Relatives	0.244	**0.589**	0.387	0.912	*–0.030*	**0.252**	0.104
(3) In-laws	0.201	**0.577**	0.460	0.837	*–0.017*	**0.179**	0.266
(7) Performance at work/housework	0.259	**0.597**	0.354	0.771	0.160	**0.159**	0.355
(8) Close friends	*–0.049*	**0.966**	0.131	0.878	*–0.019*	**0.292**	0.143
(10) Colleagues	–0.108	**0.995**	0.152	0.837	*–0.026*	**0.429**	0.114
(11) Neighbors	–0.152	**0.914**	0.341	0.734	–0.066	**0.332**	0.347
ECVf				75.8%	17.6%	5.8%	
ECVc					0.4%	0.4%	
ω	0.88	0.88		0.96	0.75	0.53	

ESEM, exploratory structural equation model; λ, factor loading; δ, uniqueness; ECVf, explained common variance of factors; ECVc, explained common variance of cross-loadings; ω, omega coefficient of composite reliability; target factor loadings on the specific factors are in bold; non-significant parameters (p ≥ 0.05) are in italics.

$p < 0.001$, RMSEA = 0.080 (90% CI 0.067–0.093), SRMR = 0.022, CFI = 0.966]. Inspection of parameter estimates (**Table 2**) indicated that the G-Factor was well-defined, with strong and positive factor loadings in all IRSS-BP items, ranging between 0.587 and 0.912 ($M_{|\lambda|}$= 0.727). Loadings of the G-Factor ranged from 0.587 to 0.657 for the intrapersonal S-Factor ($M_{|\lambda|}$= 0.626) and from 0.734 to 0.912 for the interpersonal S-Factor ($M_{|\lambda|}$= 0.828). Composite reliability estimate indicated that the G-Factor was highly reliable (ω = 0.96). The G-Factor explained about 76% of the common variance extracted. The intrapersonal S-Factor was well-defined, with significant and relevant loadings for all its items ($|\lambda|$= 0.302 to 0.650, $M_{|\lambda|}$= 0.482). Two items (item 6 and 12) loaded higher on this S-Factor than on the G-Factor, with small differences in loadings ($\Delta\lambda$ = 0.026 and 0.063 for item 6 and item 12, respectively). The remaining four items loaded more strongly on the G-Factor than the S-Factor, with differences in loadings between 0.132 (item 9) and 0.352 (item 4). The intrapersonal S-Factor showed satisfactory reliability (ω = 0.75) and explained about 18% of the variance in the items. The interpersonal S-Factor was relatively well defined ($|\lambda|$= 0.159–0.429, $M_{|\lambda|}$= 0.274), as four target items (items 2, 3, 7, and 8) had loadings lower than 0.30 (although just below 0.30 for item 8), indicating that the variance in these items was primarily used in defining the G-Factor, with differences in loadings ranging from 0.586 (item 8) to 0.660 (item 2). The remaining two items had relevant loadings but loaded higher on the G-Factor than the S-Factor ($\Delta\lambda$ = 0.408 for item 10 and 0.402 for item 12). The reliability level was acceptable (ω = 0.53), and this S-Factor explained about 6% of the common variance.

Comparison of factor loadings between the two ESEM solutions showed that the cross-loadings ranged from 0.037 to 0.346 ($M_{|\lambda|}$= 0.152) in the two-correlated factor ESEM solution, and from 0.004 to 0.160 ($M_{|\lambda|}$= 0.054) in the bifactor ESEM solution (**Table 2**). No cross-loadings > $|0.20|$ remained in the bifactor ESEM solution compared with the first-order ESEM solution, probably due to the inclusion of the G-Factor in the model. The ECV assumed by the cross-loadings in the bifactor ESEM solution was only 0.8%. The average proportion of variance in the items explained by the factors (σ^2 true total) was 63.7% in the two-correlated factor ESEM solution and 71.3% in the bifactor ESEM solution (**Supplementary Table S2**).

Overall, these results support the superiority of the bifactor model, which was retained for subsequent analyses of measurement invariance.

Measurement Invariance and Differences in Infertility-Related Stress Scale Scores Across Brazil and Italy

In the tests of measurement invariance across countries (**Table 3**), invariance of factor structure (configural), factor loadings (metric/weak), and item intercepts (scalar/strong) across Brazil and Italy was supported. Although the ΔS-B χ^2 tests were statistically significant, changes in CFI, RMSEA, and SRMR values remained acceptable across nested models. Full strict invariance was not achieved. Based on both modification indices for the strict invariance model and parameter estimates of the scalar/strong invariance model, we allowed the uniqueness of item 2 (Relatives) to be freely estimated across countries. The uniqueness of this item was 0.103 in the Brazilian sample and 0.302 in the Italian sample. The model of strict partial invariance was supported and retained in subsequent invariance tests. The models including equality constraints on the latent variance-covariance matrix and the latent factor means did not substantially decrease model fit, indicating invariant latent variance-covariance and factor means across Brazilian and Italian infertile individuals.

TABLE 3 | Goodness-of-fit indices for tests of measurement invariance (n = 1,079).

Level of invariance	*df*	S-B χ^2	Δdf	ΔS-B χ^2	CFI	ΔCFI	RMSEA	ΔRMSEA	SRMR	ΔSRMR
Configural	66	296.899*	–	–	0.964	–	0.081	–	0.022	–
Metric/weak	93	333.615**	27	50.211*	0.962	–0.002	0.069	–0.012	0.037	+0.015
Scalar/strong	102	422.639*	9	107.493**	0.950	–0.012	0.076	+0.007	0.041	+0.004
Strict	114	550.531**	12	151.742**	0.932	–0.018	0.084	+0.008	0.042	+0.001
Strict partial	113	476.290**	11	51.076*	0.943	–0.007	0.077	+0.001	0.042	+0.001
Latent variance-covariance	119	482.148**	6	8.044	0.943	0.000	0.075	–0.002	0.049	+0.007
Latent means	122	484.220**	3	1.134	0.943	0.000	0.074	–0.001	0.051	+0.002

Brazil: n = 553 (51.3%); Italy: n = 526 (48.7%); in the partial strict invariance model, uniqueness of item 2 was freely estimated across countries.
$^*p < 0.01$, $^{**}p < 0.001$.

No significant differences between Brazilian and Italian participants were found in the observed scores for the G-Factor (Brazil: M = 32.08, SD = 16.54; Italy: M = 33.00, SD = 15.54) [F(1,1077) = 0.89, p = 0.35, d = 0.06], intrapersonal S-Factor (Brazil: M = 17.57, SD = 9.06; Italy: M = 18.32, SD = 8.53) [F(1,1077) = 1.95, p = 0.16, d = 0.09] nor in the interpersonal S-Factor (Brazil: M = 14.51, SD = 8.73; Italy: M = 14.68, SD = 8.13) [F(1,1088) = 0.12, p = 0.73, d = 0.02].

Relations to Sociodemographic and Infertility-Related Characteristics and Depressive Symptomatology

Associations with sociodemographic and infertility-related characteristics and depression were calculated only in the Brazilian sample to test for the IRSS-BP concurrent validity. Age was unrelated to IRSS-BP overall (r = –0.06, p = 0.17), intrapersonal (r = –0.08, p = 0.06) and interpersonal domain scores (r = –0.03, p = 0.51). Results of group comparisons and descriptive statistics are displayed in **Table 4**. Interactions between sociodemographic and infertility-related characteristics were non-significant. Gender and duration of infertility had significant main effects on global IRSS-BP and both IRSS-BP domains. Compared to men, women reported slightly higher mean scores on global IRSS-BP scores (d = 0.36) as well as on both the intrapersonal (d = 0.36) and interpersonal (d = 0.30) domains. Participants who had been trying to conceive for 1–2 years reported lower scores than participants who had been trying for 3–4, 5–6, and >6 years in the global IRSS-BP (d = 0.40–0.50) and in the intrapersonal (d = 0.27–0.35) and interpersonal (d = 0.42–0.58) IRSS-BP domains. In the intrapersonal domain, the levels of stress reported by participants who had been trying to conceive for 1–2 years and >6 years did not differ significantly.

Repeated measures ANOVA was used to detect whether the perceived impact of infertility-related stress was greater in one of the IRSS-BP domains than in the other. Regardless of gender and duration of infertility, mean scores in the intrapersonal domain (M = 17.57, SD = 9.06) were significantly higher than scores in the interpersonal domain (M = 14.51, SD = 8.73) [F(1,545) = 93.41, p < 0.001, d = 0.34]. As shown in **Figure 1**, differences between IRSS-BP domains were primarily related to the duration of infertility. While the effect size of the difference between intrapersonal and interpersonal infertility-related stress was small in both women (d = 0.37) and men (d = 0.32), it was medium for a duration of infertility of 1–2 years (d = 0.50) and small for more than 2 years of attempting to conceive (d = 0.22–36).

In the subsample of participants who also completed the BDI-II (n = 222), 66.7% reported below-threshold and 33.3% above-threshold levels of depressive symptoms. Regardless of gender, participants with above-threshold BDI-II scores reported significantly higher scores in all IRSS-BP factors than those with below-threshold BDI-II scores (**Table 4**). The effect size of these group differences was medium-to-large for global IRSS-BP (d = 0.73), large for the intrapersonal domain (d = 0.84), and medium for the interpersonal domain (d = 0.50).

Bivariate correlations with BDI-II scores in the total subsample were r = 0.38 for global IRSS-BP, r = 0.43 for the intrapersonal domain, and r = 0.26 for the interpersonal domain (p < 0.001). Computation of correlations by gender indicated that associations with BDI-II scores among men (n = 88) were non-significant for global IRSS-BP (r = 0.18, p = 0.10) and the interpersonal domain (r = 0.08, p = 0.10), and significant and small-to-moderate for the intrapersonal domain (r = 0.24, p = 0.02). Among women (n = 134), correlations with BDI-II scores were significant and strong for total IRSS-BP and the intrapersonal domain (r = 0.48 and 0.53, respectively, p < 0.001) and moderate for the interpersonal domain (r = 0.35, p < 0.001).

DISCUSSION

The present study aimed to assess the IRSS psychometric properties in its Brazilian–Portuguese language version (IRSS-BP). We examined the factor structure, reliability, and relations to other variables of the IRSS-BP on a sample of Brazilian infertile individuals who were undergoing or seeking ART treatment. We also used a sample of infertile Italian individuals to test for measurement invariance across Brazil and Italy, as the IRSS was initially developed and validated in Italy (Casu and Gremigni, 2016).

Exploratory structural equation modeling (ESEM) with target rotation (Asparouhov and Muthén, 2009) was used in this study to investigate the IRSS-BP factor structure. We tested both the two-correlated factor model proposed in the IRSS original validation study and a bifactor model where each item simultaneously loaded on a general factor (G-Factor) and two specific factors (S-Factor intrapersonal and S-Factor

TABLE 4 | Comparisons between groups in IRSS-BP scores (n = 553).

	G-Factor			Intrapersonal domain			Interpersonal domain			
	M	***SD***	**Effect**	***M***	***SD***	**Univariate effect**	***M***	***SD***	**Univariate effect**	**Multivariate effect**
Gender			$F(1,416)$ = 9.27**			$F(1,416)$ = 9.66**			$F(1,416)$ = 6.15*	Wilks' λ = 0.98 $F(2,415)$ = 4.92**
Women (n =300)	34.75	16.75		19.06	9.18		15.69	8.88		
Men (n = 253)	28.91	15.74		15.81	8.62		13.10	8.34		
Education			$F(1,416)$ = 0.06			$F(1,416)$ = 0.04			$F(1,416)$ = 0.06	Wilks' λ = 1.00 $F(2,415)$ = 0.03
Up to high school (n = 129)	32.26	17.92		17.51	9.46		14.75	9.55		
Degree/post-degree (n = 424)	32.02	16.12		17.59	8.95		14.43	8.47		
Duration of infertility			$F(3,416)$ = 3.70*			$F(3,416)$ = 3.28*			$F(3,416)$ = 3.18*	Wilks' λ = 0.97 $F(6,830)$ = 2.16*
1—2 years (n = 188)	27.66[a]	13.72		15.74[a]	8.31		11.92[a]	6.96		
3–4 years (n = 165)	33.87[b]	17.20		18.56[b]	9.20		15.30[b]	9.01		
5-6 years (n = 121)	35.31[b]	17.36		18.77[b]	9.43		16.54[b]	9.27		
>6 years (n = 79)	33.92[b]	18.05		18.03[ab]	9.43		15.90[b]	9.69		
Infertility type			$F(1,416)$ = 1.19			$F(1,416)$ = 2.14			$F(1,416)$ = 0.27	Wilks' λ = 0.99 $F(2,415)$ = 1.30
Primary (n = 375)	32.13	16.19		17.49	8.97		14.65	8.57		
Secondary (n = 178)	31.97	17.31		17.75	9.29		14.22	9.06		
Infertility diagnosis			$F(4,416)$ = 2.55			$F(4,416)$ = 2.35			$F(4,416)$ = 2.24	Wilks' λ = 0.97 $F(8,830)$ = 1.71
Undiagnosed (n = 78)	34.12	16.95		18.87	9.46		15.24	8.63		
Male (n = 140)	29.68	13.87		16.31	7.77		13.36	7.57		
Female (n = 119)	33.08	16.98		18.40	9.38		14.68	8.90		
Both male female (n = 53)	30.09	16.90		16.40	9.67		13.70	8.57		
Unexplained (n = 163)	33.08	17.85		17.80	9.39		15.28	9.56		
BDI-II (n = 222)			$F(1,218)$ = 21.73***			$F(1,218)$ = 30.16***			$F(1,218)$ = 9.48**	Wilks' λ = 0.88 $F(2,217)$ = 15.48***
Below-thresholds (n = 148)	30.78	15.74		16.51	8.71		14.27	8.43		
Above-threshold (n = 74)	42.38	16.41		23.78	8.52		18.59	9.10		

Score range was 12–84 for global IRSS-BP (G-Factor) and 6–42 for intrapersonal and interpersonal domains (S-Factors). Means with different superscript letters in the same column significantly differ at $p < 0.05$ (Bonferroni post hoc multiple comparisons).
$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

interpersonal). The goodness of fit of the original first-order factor model was acceptable after allowing covariation between the uniqueness of two items. The two factors were strongly related, showing the same correlation of 0.72 reported in the IRSS original validation study (Casu and Gremigni, 2016). Noteworthy, such a strong correlation between the first-order factors supported the specification and testing of an alternative bifactor model, including the presence of an underlying global construct (Morin et al., 2016). The bifactor ESEM solution had an adequate fit to the data. Inspection of item loadings revealed a highly reliable G-Factor defined by positive and strong loadings for all items. All IRSS-BP items except two loaded more strongly on the G-Factor than the respective S-Factors. The G-Factor explained 76% of the common variance in the items, with the remaining 24% spread across the S-Factors. All target loadings on the intrapersonal S-Factor were significant and relevant, indicating that this S-Factor reflects meaningful specificity not represented in the G-Factor (Morin et al., 2016). Therefore, the intrapersonal S-Factor was well-defined and showed a good ω reliability estimate. The interpersonal S-Factor was less well defined, as target loadings were all significant but non-relevant for three items and just below 0.30 for one item. This indicates that the interpersonal S-Factor maintained limited specificity once the G-Factor was considered. However, the composite reliability estimate was above acceptable levels (i.e., >0.50) indicating that scores in the interpersonal S-Factor accounted for a non-negligible amount of variation beyond the G-Factor (Perreira et al., 2018).

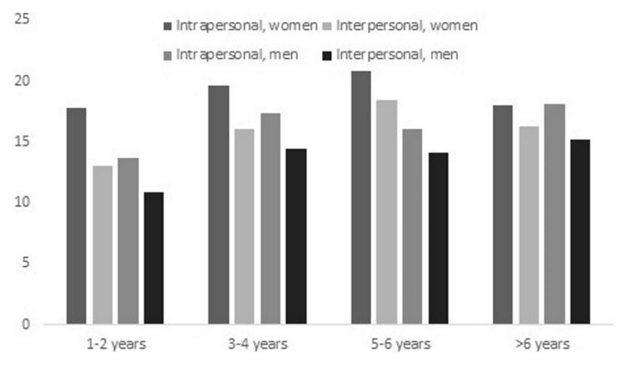

FIGURE 1 | Brazilian–Portuguese IRSS domains by gender and duration of infertility.

Comparison of ESEM solutions revealed that no cross-loadings > |0.20| remained in the bifactor solution compared with the first-order solution, supporting the presence of an underlying global construct. The average proportion of actual variance in the items explained by the factors was 71% in the bifactor solution compared to 64% in the first-order solution, indicating no decrease in reliability in the bifactor solution. Based on this, and considering the presence of a well-defined, highly reliable G-Factor and S-Factors retaining at least some specificity and acceptable levels of composite reliability, we considered that the bifactor model provided a better representation of the data (Morin et al., 2016). Noteworthy, the limited specificity of the interpersonal S-Factor when the variance explained by the G-Factor is considered supports the need for a bifactor representation of the data (Sánchez-Oliva et al., 2017). Altogether, the G-Factor provides a direct estimate of global infertility-related stress based on responses to all items. In contrast, the S-Factors represent specific components of infertility-related stress not already explained by the global component and unique to the intrapersonal and interpersonal subsets of items.

The bifactor model replicated well across Brazil and Italy, and multigroup analyses supported strict measurement invariance. Invariance of latent variance-covariance and latent means was also observed, indicating similar levels of inter-individual variability and latent factor scores across countries. Noteworthy, evidence of measurement invariance was established despite some differences in demographic (age and education) and infertility-related characteristics (infertility type and diagnosis) across Brazilian and Italian participants. Invariance of the IRSS bifactor model suggests that the coexistence of global infertility-related stress and its components represents a common pattern across Brazilian and Italian cultures. Brazilian and Italian infertile individuals thus seem to conceptualize and evaluate the burden of infertility in the same way (Steenkamp and Baumgartner, 1998). It is in line with the notion that infertility is a stress-triggering factor (Yazdani et al., 2017) that can affect the well-being and relationships of infertile people regardless of cultural differences (Stulhofer et al., 2013; Qadir et al., 2015). Brazilian and Italian participants in this study reported similar levels of infertility-related stress, both overall and in the intrapersonal and interpersonal domains. A previous study using a different self-report measure of infertility-related stress also reported no differences in mean scores between Brazilian and Italian infertile individuals (Gremigni et al., 2018).

As for the relations of IRSS-BP to sociodemographic and infertility-related characteristics, age, education, infertility type, and infertility diagnosis were unrelated to levels of infertility-related stress, in line with the IRSS original validation study (Casu and Gremigni, 2016) and other research in the Brazilian context (Casu et al., 2018, 2019), although contrary to other studies (Patel et al., 2016; Lakatos et al., 2017; Zaidouni et al., 2018). We found instead that IRSS-BP scores varied depending on gender and duration of infertility. Women showed slightly higher infertility-related stress than men in global IRSS-BP and intrapersonal and interpersonal domains. This finding is in line with previous research and literature reviews reporting that infertility is more distressing for women than men, and women experience greater infertility-related stress than their male counterparts (Greil et al., 2010; Ying et al., 2015; Chaves et al., 2019). It is likely due to cultural stereotypes and gender role expectations, which emphasize the centrality of motherhood and childbearing in women's social function, especially in developing societies (Greil et al., 2010; Ying et al., 2015). Previous Brazilian studies using the IRSS also reported a link between female gender and higher global infertility-related stress (Casu et al., 2018, 2019). In the original validation study of the IRSS, Italian women reported a significantly higher level of infertility stress than men in the intrapersonal but not in the interpersonal domain (Casu and Gremigni, 2016). This discrepancy might be due to different sample characteristics. While participants in this study were seeking or already undergoing fertility treatment at the time of data collection, those in the Italian validation study were at their first visit to the fertility clinic. A gender gap has been observed in the diagnostic journey of the infertile couple, where the woman is the recipient of most diagnostic examinations (Gullo et al., 2021). Therefore, it is plausible that women in the Italian validation study were experiencing an acute stress reaction that characterizes initial stages involving the diagnostic work-up and mainly affects the intrapersonal life domain (Berg and Wilson, 1991).

We found that as the duration of the infertility problem increased, so did infertility-related stress. However, after more than 6 years of attempting conception, the levels of infertility-related stress tended to decrease, and there were no significant differences in the reported impact of infertility in the intrapersonal domain between individuals who had been trying to conceive for 1–2 years and those with more than 6 years of infertility. This is coherent with research indicating that a time frame of 3–6 years of infertility is characterized by the most significant risk for emotional maladjustment (Domar et al., 1992; Drosdzol and Skrzypulec, 2009). As previously suggested, during the first 1–2 years, infertile individuals tend to be optimistic about the possibility of conception, then begin to feel hopeless, and finally start to solve their feelings and accept remaining childless or turn to alternative parenthood options such as adoption (Domar et al., 1992; Adewunmi et al., 2012).

Our results also indicated that the perceived impact of infertility was significantly higher in the intrapersonal than in the interpersonal domain, as also reported in the IRSS original validation study (Casu and Gremigni, 2016) and coherent with previous evidence that the stressful implications of infertility concern the domain of self, more than social life (Karabulut et al., 2013; Swanson and Braverman, 2021). In this study, the difference between levels of infertility-related stress in the two life domains was moderate for a duration of infertility of 1–2 years and small for more than 2 years of infertility duration. Specifically, scores in the interpersonal domain were higher as the duration of infertility increased. This is coherent with previous findings that a longer duration of infertility was associated with greater infertility-related stress in the social areas of family and social relationships and work-life (Karabulut et al., 2013; Keramat et al., 2014;

Santoro et al., 2016). However, the fertility treatment stage and number of previous treatment cycles might also be associated with infertility-related stress changes in both domains (Volpini et al., 2020). Therefore, prospective studies should be conducted to examine whether changes in IRSS-BP scores are a response to the duration of infertility or the duration of infertility treatments.

As for relations to depressive symptomatology, regardless of gender, participants with above-threshold levels of depression showed higher total and subscale IRSS-BP scores than participants with below-threshold depression. Differences among the BDI-II groups were large in the intrapersonal domain, medium-to-large in global infertility-related stress, and moderate in the interpersonal domain. These results are coherent with those reported in the Italian validation study (Casu and Gremigni, 2016), where highly depressed women and men had significantly higher scores than non-depressed ones in both IRSS components, with a more substantial effect in the intrapersonal than the interpersonal domain. In this study, bivariate correlations with BDI-II scores for global IRSS-BP and the interpersonal domain were strong and moderate, respectively, among women but non-significant among men. In contrast, those for the intrapersonal domain were strong among women and small-to-moderate among men. Differently, in the Italian validation study, the correlations of BDI-II with global and intrapersonal infertility-related stress were strong for men and moderate for women, and those with the interpersonal IRSS domain were moderate for both genders. Such a discrepancy might be due to cultural differences, as Brazil has a more distinct patriarchal culture and collectivist features than Italy (Oyserman and Sorensen, 2009). Previous research on Brazilian infertile individuals indeed indicated that Brazilian women, compared to their male counterparts, had more negative emotional reactions to interpersonal aspects such as being questioned about childlessness by relatives and friends and being invited to a child's birthday party (Franco et al., 2002). Also, the correlations observed in this study are coherent with evidence that the associations of depression with infertility-related stress were stronger for women than for men (Chaves et al., 2019), and those with stress in the social areas of life were significant for women but not for men (Peterson et al., 2003), thus supporting the concurrent validity of the IRSS-BP.

Limitations and Future Directions

The present study has some limitations. First, using a convenience sample and the collection of Brazilian data in one clinical setting only limit the generalizability of results. Second, reliability was assessed as model-based internal consistency only; future studies should therefore assess test–retest stability as well. Third, the performance of the IRSS-BP was not compared with other measures of the same construct, such as the FPI (Newton et al., 1999) and the FertiQoL (Boivin et al., 2011), and associations with other psychological variables were limited to levels of depression. Due to the small size of the subsample that completed the BDI-II, we could not examine the IRSS-BP nomological validity by integrating a BDI-II latent factor in the bifactor ESEM measurement model. This would allow for a simultaneous examination of the associations of global and specific IRSS-BP factors with the external criterion variables to clarify whether the IRSS-BP S-Factors have sufficient specificity to result in significant relations with the criteria over and above the prediction provided by the G-Factor. Future studies with larger samples should test for the validity of the IRSS-BP bifactor model within a semantic network of similar or different constructs. Fourth, other individual characteristics and infertility conditions that we have not measured could influence the individual's infertility-related stress; therefore, a more extensive collection of sociodemographic and clinical information should be done in future research. Finally, the sample was composed of independent individuals, although infertility-related stress is a shared experience within a couple (Pasch and Sullivan, 2017). Thus, future studies could collect data from both couple members and use a dyadic approach to test for dyadic invariance of the IRSS-BP bifactor model across partners of the infertile couple (Claxton et al., 2015).

CONCLUSION

Our findings indicated that the underlying structure of IRSS-BP scores was best represented by a bifactor solution incorporating an overarching infertility-related stress factor and two specific components of intrapersonal and interpersonal life domains affected by infertility stress. Both the general and the specific IRSS-BP factors showed adequate levels of composite reliability and validity evidence based on relations to sociodemographic and infertility-related characteristics and depressive symptomatology. Therefore, when using the IRSS-BP in research and clinical practice, we suggest considering the total score and the two subscale scores of intrapersonal and interpersonal infertility-related stress.

Altogether, the present study provides initial evidence of validity and reliability for the Brazilian-Portuguese language version of the IRSS to be used in the Brazilian context to rapidly assess the burden of infertility at both global and domain-specific levels. The IRSS-BP could be used by staff in fertility clinics to identify patients who need support to deal with the stressful impact of infertility on the areas of the self and social life, as recommended by international guidelines for routine psychosocial care as a means to improve infertile patients' well-being and compliance with treatment (Gameiro et al., 2015).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article are available on request from the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Faculdade de Medicina do

ABC and Ethics Committee of University of Bologna. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GC: data analysis and interpretation, software, visualization, writing-original draft, and writing-review and editing. VZ: study concept and design, provision of study materials and patients, collection and assembly of data, data analysis, writing-original draft, and writing-review and editing. AP: study concept and design, provision of study materials and patients, writing review and editing. EM and BB: collection and assembly of data, and writing-original draft. CB: writing-original draft. PG: study concept and design, provision of study materials, supervision, and writing–review and editing. All authors contributed to the manuscript and approved the final version.

FUNDING

This research was supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) of the São Paulo State, Brazil. Number: 2019/17853-2.

ACKNOWLEDGMENTS

Thanks to Maria do Carmo Fernandes Martins for her contributions to the improvement of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.784222/full#supplementary-material

REFERENCES

Adewunmi, A. A., Etti, E. A., Tayo, A. O., Rabiu, K. A., Akindele, R. A., Ottun, T. A., et al. (2012). Factors associated with acceptability of child adoption as a management option for infertility among women in a developing country. *Int. J. Womens Health* 4, 365–372. doi: 10.2147/IJWH.S31598

Anaman-Torgbor, J. A., Jonathan, J. W. A., Asare, L., Osarfo, B., Attivor, R., Bonsu, A., et al. (2021). Experiences of women undergoing assisted reproductive technology in Ghana: a qualitative analysis of their experiences. *PLoS One* 16:e0255957. doi: 10.1371/journal.pone.0255957

Arya, S. T., and Dibb, B. (2016). The experience of infertility treatment: the male perspective. *Hum. Fertil.* 19, 242–248. doi: 10.1080/14647273.2016.1222083

Asparouhov, T., and Muthén, B. (2009). Exploratory structural equation modeling. *Struct. Equ. Model.* 16, 397–438. doi: 10.1080/10705510903008204

Beck, A. T., Steer, R. A., and Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.

Berg, B. J., and Wilson, J. F. (1991). Psychological functioning across stages of treatment for infertility. *J. Behav. Med.* 14, 11–26. doi: 10.1007/BF00844765

Boivin, J., Bunting, L., Collins, J. A., and Nygren, K. G. (2007). International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Hum. Reprod.* 22, 1506–1512. doi: 10.1093/humrep/dem046

Boivin, J., Takefman, J., and Braverman, A. (2011). The fertility quality of life (FertiQoL) tool: development and general psychometric properties. *Fertil. Steril.* 96, 409–415. doi: 10.1093/humrep/der171

Casu, G., and Gremigni, P. (2016). Screening for infertility-related stress at the time of initial infertility consultation: psychometric properties of a brief measure. *J. Adv. Nurs.* 72, 693–706. doi: 10.1111/jan.12830

Casu, G., Ulivi, G., Zaia, V., Fernandes Martins, M. D. C., Parente Barbosa, C., and Gremigni, P. (2018). Spirituality, infertility-related stress, and quality of life in Brazilian infertile couples: analysis using the actor-partner interdependence mediation model. *Res. Nurs. Health* 41, 156–165. doi: 10.1002/nur.21860

Casu, G., Zaia, V., Fernandes Martins, M. D. C., Parente Barbosa, C., and Gremigni, P. (2019). A dyadic mediation study on social support, coping, and stress among couples starting fertility treatment. *J. Fam. Psychol.* 33, 315–326. doi: 10.1037/fam0000502

Chaves, C., Canavarro, M. C., and Moura-Ramos, M. (2019). The role of dyadic coping on the marital and emotional adjustment of couples with infertility. *Fam. Process* 58, 509–523. doi: 10.1111/famp.12364

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Model.* 14, 464–504. doi: 10.1080/10705510701301834

Claxton, S. E., Deluca, H. K., and van Dulmen, M. H. (2015). Testing psychometric properties in dyadic data using confirmatory factor analysis: current practices and recommendations. *TPM Test. Psychom. Methodol. Appl. Psychol.* 22, 181–198. doi: 10.4473/TPM22.2.2

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. New Jersey, NJ: Lawrence Erlbaum.

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *J. Exp. Soc. Psychol.* 66, 4–19. doi: 10.1016/j.jesp.2015.07.006

Domar, A. D., Broome, A., Zuttermeister, P. C., Seibel, M., and Friedman, R. (1992). The prevalence and predictability of depression in infertile women. *Fertil. Steril.* 58, 1158–1163. doi: 10.1016/S0015-0282(16)55562-9

Domar, A. D., Rooney, K., Hacker, M. R., Sakkas, D., and Dodge, L. E. (2018). Burden of care is the primary reason why insured women terminate in vitro fertilization treatment. *Fertil. Steril.* 109, 1121–1126. doi: 10.1016/j.fertnstert.2018.02.130

Drosdzol, A., and Skrzypulec, V. (2009). Depression and anxiety among Polish infertile couples – an evaluative prevalence study. *J. Psychosom. Obstet. Gynecol.* 30, 11–20. doi: 10.1080/01674820902830276

Franco, J. G., Baruffi, R. L. R., Mauri, A. L., Petersen, C. G., Felipe, V., and Garbellini, E. (2002). Psychological evaluation test for infertile couples. *J. Assist. Reprod. Genet.* 19, 269–273. doi: 10.1023/A:1015706829102

Gameiro, S., Boivin, J., Dancet, E., de Klerk, C., Emery, M., Lewis-Jones, C., et al. (2015). ESHRE guideline: routine psychosocial care in infertility and medically assisted reproduction—a guide for fertility staff. *Hum. Reprod.* 30, 2476–2485. doi: 10.1093/humrep/dev177

Gomes-Oliveira, M. H., Gorenstein, C., Lotufo Neto, F., Andrade, L. H., and Wang, Y. P. (2012). Validation of the brazilian portuguese version of the beck depression Inventory-II in a community sample. *Braz. J. Psychiatry* 34, 389–394. doi: 10.1016/j.rbp.2012.03.005

Greil, A. L., Slauson-Blevins, K., and McQuillan, J. (2010). The experience of infertility: a review of recent literature. *Sociol. Health Illn.* 32, 140–162. doi: 10.1111/j.1467-9566.2009.01213.x

Gremigni, P., Casu, G., Mantoani Zaia, V., Viana Heleno, M. G., Conversano, C., and Barbosa, C. P. (2018). Sexual satisfaction among involuntarily childless women: a cross-cultural study in Italy and Brazil. *Women Health* 58, 1–15. doi: 10.1080/03630242.2016.1267690

Gullo, G., Cucinella, G., Perino, A., Gullo, D., Segreto, D., Laganà, A. S., et al. (2021). The gender gap in the diagnostic-therapeutic journey of the infertile couple. *Int. J. Environ. Res. Public Health* 18:6184. doi: 10.3390/ijerph18126184

Ha, J. Y., and Ban, S. H. (2020). Effect of resilience on infertile couples' quality of life: an actor-partner interdependence model approach. *Health Qual. Life Outcomes* 18:295. doi: 10.1186/s12955-020-01550-6

Hasanpoor-Azghdy, S. B., Simbar, M., and Vedadhir, A. (2015). The social consequences of infertility among Iranian women: a qualitative study. *Int. J. Fertil. Steril.* 8, 409–420. doi: 10.22074/ijfs.2015.4181

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Instituto Brasileiro de Geografia e Estatística [IBGE] (2010). *Censo Demográfico da População Brasileira*. https://www.ibge.gov.br/ (accessed September 3, 2021).

Karabulut, A., Özkan, S., and Oğuz, N. (2013). Predictors of fertility quality of life (FertiQoL) in infertile women: analysis of confounding factors. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 170, 193–197. doi: 10.1016/j.ejogrb.2013.06.029

Keramat, A., Masoumi, S. Z., Mousavi, S. A., Poorolajal, J., Shobeiri, F., and Hazavehie, S. M. M. (2014). Quality of life and its related factors in infertile couples. *J. Res. Health Sci.* 14, 57–64.

Kiani, Z., Simbar, M., Hajian, S., and Zayeri, F. (2021). The prevalence of depression symptoms among infertile women: a systematic review and meta-analysis. *Fertil. Res. Pract.* 7:6. doi: 10.1186/s40738-021-00098-3

Kline, R. (2005). *Principles and Practice of Structural Equation Modeling*, 2nd Edn. New York, NY: The Guilford Press.

Lakatos, E., Szigeti, J. F., Ujma, P. P., Sexty, R., and Balog, P. (2017). Anxiety and depression among infertile women: a cross-sectional survey from Hungary. *BMC Women's Health* 17:48. doi: 10.1186/s12905-017-0410-2

Loke, A. Y., Yu, P. L., and Hayter, M. (2012). Experiences of sub-fertility among Chinese couples in Hong Kong: a qualitative study. *J. Clin. Nurs.* 21, 504–512. doi: 10.1111/j.1365-2702.2010.03632.x

Luk, B. H., and Loke, A. Y. (2015). The impact of infertility on the psychological well-being, marital relationships, sexual relationships, and quality of life of couples: a systematic review. *J. Sex Marital Ther.* 41, 610–625. doi: 10.1080/0092623X.2014.958789

Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., and Greenglass, E. (2015). The inter-item standard deviation (ISD): an index that discriminates between conscientious and random responders. *Pers. Indiv. Differ.* 84, 79–83. doi: 10.1016/j.paid.2014.08.021

Marsh, H. W., Morin, A. J., Parker, P. D., and Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annu. Rev. Clin. Psychol.* 10, 85–110. doi: 10.1146/annurev-clinpsy-032813-153700

Mascarenhas, M. N., Flaxman, S. R., Boerma, T., Vanderpoel, S., and Stevens, G. A. (2012). National, regional, and global trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. *PLoS Med.* 9:e1001356. doi: 10.1371/journal.pmed.1001356

Mayring, P. (2014). *Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution*. Klagenfurt: GESIS.

McDonald, R. P. (1970). Theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *Brit. J. Math. Stat. Psychol.* 23, 1–21. doi: 10.1111/j.2044-8317.1970.tb00432.x

Morin, A. J. S., Arens, A., and Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Struct. Equ. Model.* 23, 116–139. doi: 10.1080/10705511.2014.961800

Morin, A. J. S., Marsh, H. W., and Nagengast, B. (2013). "Exploratory structural equation modeling," in *Structural Equation Modeling: A Second Course*, eds G. R. Hancock and R. O. Mueller (Charlotte, NC: Information Age), 395–436.

Morin, A. J. S., Myers, N. D., and Lee, S. (2020). "Modern factor analytic techniques: bifactor models, exploratory structural equation modeling (ESEM) and bifactor-ESEM," in *Handbook of Sport Psychology*, 4th Edn, eds G. Tenenbaum and R. C. Eklund (London: Wiley), 1044–1073.

Muthén, L. K., and Muthén, B. O. (1998–2017). *Mplus User's Guide*, 8th Edn. Los Angeles, CA: Muthén and Muthén.

Newton, C. R., Sherrard, W., and Glavac, I. (1999). The Fertility Problem Inventory: measuring perceived infertility-related stress. *Fertil. Steril.* 72, 54–62. doi: 10.1016/S0015-0282(99)00164-8

Ombelet, W., Cooke, I., Dyer, S., Serour, G., and Devroey, P. (2008). Infertility and the provision of infertility medical services in developing countries. *Hum. Reprod. Update* 14, 605–621. doi: 10.1093/humupd/dmn042

Oyserman, D., and Sorensen, N. (2009). "Understanding cultural syndrome effects on what and how we think: a situated cognition model," in *Understanding Culture: Theory, Research and Application*, eds R. Wyer, Y. Y. Hong, and C. Y. Chiu (New York, NY: Psychology Press), 25–52.

Öztürk, R., Bloom, T. L., Li, Y., and Bullock, L. F. (2021a). Stress, stigma, violence experiences and social support of US infertile women. *J. Reprod. Infant Psychol.* 39, 205–217. doi: 10.1080/02646838.2020.1754373

Öztürk, R., Herbell, K., Morton, J., and Bloom, T. (2021b). "The worst time of my life": treatment-related stress and unmet needs of women living with infertility. *Am. J. Community Psychol.* 49, 1121–1133. doi: 10.1002/jcop.22527

Pasch, L. A., and Sullivan, K. T. (2017). Stress and coping in couples facing infertility. *Curr. Opin. Psychol.* 13, 131–135. doi: 10.1016/j.copsyc.2016.07.004

Patel, A., Sharma, P. S. V. N., Kumar, P., and Binu, V. S. (2018). Illness cognitions, anxiety, and depression in men and women undergoing fertility treatments: a dyadic approach. *J. Hum. Reprod. Sci.* 11, 180–189. doi: 10.4103/jhrs.JHRS_119_17

Patel, A., Sharma, P. S. V. N., Narayan, P., Binu, V. S., Dinesh, N., and Pai, P. J. (2016). Prevalence and predictors of infertility-specific stress in women diagnosed with primary infertility: a clinic-based study. *J. Hum. Reprod. Sci.* 9, 28–34. doi: 10.4103/0974-1208.178630

Pedro, J., Sobral, M. P., Mesquita-Guimaraes, J., Leal, C., Costa, M. E., and Martins, M. V. (2017). Couples' discontinuation of fertility treatments: a longitudinal study on demographic, biomedical, and psychosocial risk factors. *J. Assist. Reprod. Genet.* 34, 217–224. doi: 10.1007/s10815-016-0844-8

Perreira, T. A., Morin, A. J. S., Hebert, M., Gillet, N., Houle, S. A., and Berta, W. (2018). The short form of the Workplace Affective Commitment Multidimensional Questionnaire (WACMQ-S): a bifactor-ESEM approach among healthcare professionals. *J. Vocat. Behav.* 106, 62–83. doi: 10.1016/j.jvb.2017.12.004

Peterson, B. D., Newton, C. R., and Rosen, K. H. (2003). Examining congruence between partners' perceived infertility-related stress and its relationship to marital adjustment and depression in infertile couples. *Fam. Process* 42, 59–70. doi: 10.1111/j.1545-5300.2003.00059.x

Qadir, F., Khalid, A., and Medhin, G. (2015). Social support, marital adjustment, and psychological distress among women with primary infertility in Pakistan. *Women Health* 55, 432–446. doi: 10.1080/03630242.2015.1022687

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behav. Res.* 47, 667–696. doi: 10.1080/00273171.2012.715555

Reise, S. P., Scheines, R., Widaman, K. F., and Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: a bifactor perspective. *Educ. Psychol. Meas.* 73, 5–26. doi: 10.1177/0013164412449831

Rhemtulla, M., Brosseau-Liard, P. E., and Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* 17, 354–373. doi: 10.1037/a0029315

Rockliff, H. E., Lightman, S. L., Rhidian, E., Buchanan, H., Gordon, U., and Vedhara, K. (2014). A systematic review of psychosocial factors associated with emotional adjustment in in vitro fertilization patients. *Hum. Reprod. Update* 20, 594–613. doi: 10.1093/humupd/dmu010

Rooney, K. L., and Domar, A. D. (2016). The impact of stress on fertility treatment. *Curr. Opin. Obstet. Gynecol.* 28, 198–201. doi: 10.1097/GCO.0000000000000261

Sánchez-Oliva, D., Morin, A. J., Teixeira, P. J., Carraça, E. V., Palmeira, A. L., and Silva, M. N. (2017). A bifactor exploratory structural equation modeling representation of the structure of the basic psychological needs at work scale. *J. Vocat. Behav.* 98, 173–187. doi: 10.1016/j.jvb.2016.12.001

Santoro, N., Eisenberg, E., Trussell, J. C., Craig, L. B., Gracia, C., Huang, H., et al. (2016). Fertility-related quality of life from two RCT cohorts with infertility: unexplained infertility and polycystic ovary syndrome. *Hum. Reprod.* 31, 2268–2279. doi: 10.1093/humrep/dew175

Schmidt, L. (2009). Social and psychological consequences of infertility and assisted reproduction–what are the research priorities? *Hum. Fertil.* 12, 14–20. doi: 10.1080/14647270802331487

Schmidt, L., Holstein, B. E., Christensen, U., and Boivin, J. (2005). Communication and coping as predictors of fertility problem stress: cohort study of 816 participants who did not achieve a delivery after 12 months of fertility treatment. *Hum. Reprod.* 20, 3248–3256. doi: 10.1093/humrep/dei193

Segerstrom, S. C., and Miller, G. E. (2004). Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychol. Bull.* 130, 601–630. doi: 10.1037/0033-2909.130.4.601

Steenkamp, J. B. E. M., and Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *J. Consum. Res.* 25, 78–90. doi: 10.1086/209528

Stulhofer, A., Traeen, B., and Carvalheira, A. (2013). Job-related strain and sexual health difficulties among heterosexual men from three European countries: the role of culture and emotional support. *J. Sex Med.* 10, 747–756. doi: 10.1111/j.1743-6109.2012.02967.x

Swanson, A., and Braverman, A. M. (2021). Psychological components of infertility. *Fam. Court Rev.* 59, 67–82. doi: 10.1111/fcre.12552

Volpini, L., Mazza, C., Mallia, L., Guglielmino, N., Rossi Berluti, F., Fernandes, M., et al. (2020). Psychometric properties of the FertiQoL questionnaire in Italian infertile women in different stages of treatment. *J. Reprod. Infant Psychol.* 38, 324–339. doi: 10.1080/02646838.2019.1698017

World Health Organization [WHO] (2006). *Reproductive Health Indicators: Guidelines for Their Generation, Interpretation and Analysis for Global Monitoring*. https://apps.who.int/iris/bitstream/handle/10665/43185/924156315X_eng.pdf (accessed September 3, 2021).

Yazdani, F., Elyasi, F., Peyvandi, S., Moosazadeh, M., Galekolaee, K. S., Kalantari, F., et al. (2017). Counseling-supportive interventions to decrease infertile women's perceived stress: a systematic review. *Electron. Physician* 9, 4694–4702. doi: 10.19082/4694

Ying, L. Y., Wu, L. H., and Loke, A. Y. (2015). Gender differences in experiences with and adjustments to infertility: a literature review. *Int. J. Nurs. Stud.* 52, 1640–1652. doi: 10.1016/j.ijnurstu.2015.05.004

Zaidouni, A., Fatima, O., Amal, B., Siham, A., Houyam, H., Jalal, K., et al. (2018). Predictors of infertility stress among couples diagnosed in a public center for assisted reproductive technology. *J. Hum. Reprod. Sci.* 11, 376–383. doi: 10.4103/jhrs.JHRS_93_18

Zegers-Hochschild, F., Adamson, G. D., Dyer, S., Racowsky, C., de Mouzon, J., Sokol, R., et al. (2017). The international glossary on infertility and fertility care. *Hum. Reprod.* 32, 1786–1801. doi: 10.1093/humrep/dex234

Ziegler, M., Kemper, C. J., and Kruyen, P. (2014). Short scales – Five misunderstandings and ways to overcome them. *J. Individ. Differ.* 35, 185–189.

Zurlo, M. C., Cattaneo Della Volta, M. F., and Vallone, F. (2020). Infertility-related stress and psychological health outcomes in infertile couples undergoing medical treatments: testing a multi-dimensional model. *J. Clin. Psychol. Med. Settings* 27, 662–676. doi: 10.1007/s10880-019-09653-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 01 February 2022
doi: 10.3389/fpsyg.2022.799967

The Chinese Inventory of Psychosocial Balance Short-Form Questionnaire for the Older Adults: Validity and Reliability Study

Pei-Yun Chen[1], Wen-Chao Ho[1]*, Chyi Lo[2,3]* and Tzu-Pei Yeh[2,3]**

[1] Department of Public Health, China Medical University (Taiwan), Taichung, Taiwan, [2] School of Nursing, China Medical University (Taiwan), Taichung, Taiwan, [3] China Medical University Hospital, Taichung, Taiwan

OPEN ACCESS

Edited by:
Paola Gremigni,
University of Bologna, Italy

Reviewed by:
Salman Shahzad,
University of Karachi, Pakistan
Dimitrios Stamovlasis,
Aristotle University of Thessaloniki, Greece

****Correspondence:***
Pei-Yun Chen
peiyun0203@gmail.com
Wen-Chao Ho
wcho@mail.cmu.edu.tw
Chyi Lo
chyilo@mail.cmu.edu.tw
Tzu-Pei Yeh
tzupeiyeh@mail.cmu.edu.tw

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *22 October 2021*
Accepted: *04 January 2022*
Published: *01 February 2022*

Citation:
Chen P-Y, Ho W-C, Lo C and Yeh T-P (2022) The Chinese Inventory of Psychosocial Balance Short-Form Questionnaire for the Older Adults: Validity and Reliability Study. Front. Psychol. 13:799967. doi: 10.3389/fpsyg.2022.799967

Background: Drawing from Erikson's theory, Domino and Affonso constructed the Inventory of Psychosocial Balance (IPB), a scale with satisfactory reliability and validity. However, the lack of a credible Chinese version of the scale may hinder research on ego development in Taiwan. The aim of the present study was to construct a short form Chinese IPB. In addition, factor analysis was employed to shorten the original 120-item scale to make it suitable for application in the older adults in the future.

Methods: The study involved three steps: The first step was to establish the 120-items of the Chinese Inventory of Psychosocial Balance (C-IPB), and we conducted translation, back-translation, expert validity, and reliability of pilot study for this step. Following the first step was to construct the short-form C-IPB (CIPB-SF) in the second step, and the CIPB-SF was developed *via* item analysis and factor analysis. Finally, we assessed the reliability and validity of the CIPB-SF *via* structural equation model in the third step.

Results: Three hundred eight older adults without cognitive disorder completed the IPB. The 40-item CIPB-SF was completed through item analysis and factor analysis. The internal consistency test of CIPB-SF and the eight stages were good (Cronbach's $\alpha = 0.81$–0.89). The CIPB-SF had acceptable validity, except in the intimacy and identity stages, in which validity was only fair. Compared with the IPB, the CIPB-SF had good reliability and acceptable validity. However, because of its conciseness, the 40-item CIPB-SF was more suited for application among the Chinese elderly population because its application avoids physical overload.

Conclusion: The CIPB-SF served as a concise scale for assessing ego development in our study. This scale can also serve as a useful tool for convenient screening in the future.

Keywords: ego, life span, validity, reliability, elderly

INTRODUCTION

The advancement of medical progression increases human longevity, but this does not make death avoidable. Adults' awareness of death may be the source of anxiousness and stress, especially in the last stages of life in older adults. Accompanying degradation of physiology, the older adults could be aware that their death is approaching, and so appear some negative emotions such as depression (Busch et al., 2018). According to the Erikson theory, while older people in the last stage of life may accept their past and find meanings from their lives, they may reach ego integrity; which means they may face death peacefully (Erikson, 1963).

The psychosocial development in older adults is influenced by their past life experiences (Ardelt et al., 2018). Erikson's life stage theory of psychosocial development is a rare psychosocial theory that encompasses the entire life span and encompasses eight age-graded stages of ego development, ranging from infancy to late adulthood. According to Erikson's theory, failure to successfully complete a stage can result in reduce ability to complete subsequent stages, in turn resulting in an unhealthy personality and a sense of self (Erikson, 1963, 1980, 1982). A specific type of crisis occurs in each stage, during which an individual cannot achieve the goal set for each stage (Erikson, 1963, 1982). The task–crisis pairs are as follows: trust vs. mistrust, autonomy vs. shame and doubt, initiative vs. guilt, industry vs. inferiority, identity vs. role confusion, intimacy vs. isolation, generativity vs. stagnation, and integrity vs. despair. The final stage represents the crisis of "integrity vs. despair," which can be described as the process through which older adults try to find meanings in their lives. Erikson proposed that individuals who had been unsuccessful in resolving earlier crises would face difficulties in resolving later psychosocial crises as well (Erikson, 1963). An individual's personality subconsciously affects their psychosocial functioning and mental health (Dezutter et al., 2016). Early psychosocial crises may worsen an individual's depression, self-neglect tendencies, and morbidity. Furthermore, early psychosocial crises may negatively influence the well-being of the older adults (Cuijpers et al., 2013; Saint Onge et al., 2014). On the contrary, by improving this, ego-integrity can be achieved successfully aging and a better quality of life (QOL; Min, 2020). The ego integrity in older people is important; however, a valid measure tool is deficient in Taiwan; the evaluation of ego integrity and related research in older adults is limited.

The measurement tools which include "The Northwestern ego-integrity scale (NEIS)" (Westerhof et al., 2017), "ego-integrity/despair scale" "(Van Hiel and Vansteenkiste, 2009)," and the "inventory of psychosocial balance (IPB)" "(Domino and Affonso, 1990)"were developed according to Erikson theory and comprehensively used in ego integrity in older adults. NEIS includes two subscales (ego-despair and ego-integrity) with 15 items. The higher item mean scores indicate more despair and ego-integrity, respectively. The Cronbach α of NEIS is 0.74 in ego integrity and 0.75 in despair (Westerhof et al., 2017), and it has been applied in many studies (Kleijn et al., 2016; Westerhof et al., 2017).

Ego-integrity/despair scale, developed by Van Hiel and Vansteenkiste (2009), is a 5-point Likert scale, the scores range from 1 (Completely not true) to 5 (Completely true). This scale included ego-despair and ego-integrity sub-scale with 3 items, respectively. The higher mean scores indicated more despair and ego-integrity, and the Cronbach's alphas were 0.80 in ego integrity and 0.85 in despair (Van Hiel and Vansteenkiste, 2009); it has been used in several studies (Dezutter et al., 2016; Derdaele et al., 2019).

Drawing from Erikson's theory, Domino and Affonso (1990) constructed the IPB, which has good reliability (Cronbach's α ranging from 0.64 to 0.79.) and validity and has been applied in many studies (Hannah et al., 1996; Brennan and MacMillan, 2006; Beaumont and Pratt, 2011). IPB covers eight stages of Ericson theory, and each stage contains 15 items (120 items in total); the items were responded in 5-point Likert scale. The higher scores represent better self-development in certain stage (Domino and Affonso, 1990).

Although these three measurement tools all possess good validity and reliability in many studies, a credible Chinese version of these tools in ego development is lacking in Taiwan. Therefore, developing a Chinese version of the scale for ego development measurement is necessary.

The NEIS and ego-integrity/despair scale only include the older adult stage, whereas the IPB covers all eight stages of self-development. The Erikson theory is an important and rare psychological theory because it involves all stages in life, which is valid in older adult mental development. The IPB could be applied in various research more comprehensively; therefore, this research selected IPB to develop a Chinese tool. However, the original version contains 120 items which may be difficult for older adults to complete the questionnaire. Therefore, developing a short version IPB in Chinese is considered cautious and necessary. The aim of this study was to develop a short version Chinese IPB.

STUDY MEASURES

Survey tools in this study included sociodemographic characteristics (age, gender, marital status, religious, self-perceived economic situation, and educational level), IPB, and MOS 36-item short-form health survey (SF-36).

Inventory of Psychosocial Balance

The IPB is a scale that draws from Erikson's ego development theory (Domino and Affonso, 1990). This scale initially had 346 items determined with reference to relevant studies and was developed to reflect both positive and negative aspects of the eight stages. The responses are recorded on a 5-point Likert scale, ranging from strongly agree to strongly disagree, and are scored between 1 and 5 points. Five psychologists familiar with Erikson's theory reviewed each item for clarity of meaning and identified each item's relevance to a specific stage, after which they wrote written descriptions of each life stage, summarized on the basis of Erikson's writings. A total of 208 of the original 346 items survived this first step (Domino and Affonso, 1990).

A total of 528 subjects, including students, adults, older adults, and retired adults, were included, and their age ranged from 21 to 88 years. The second process was factor analysis. The factor loading had to be at least 0.3, and the items had to significantly ($p < 0.01$) correlate with the appropriate self-rating. Finally, the 15 best items were selected (Domino and Affonso, 1990). The IPB thus obtained had good reliability demonstrated through internal consistency, evident in a Cronbach's α ranging from 0.64 to 0.79. The 4–5 week test–retest reliability ranged from 0.79 to 0.90 (Hannah et al., 1996). The validation of the IPB was compared with that of the California Psychological Inventory (CPI)–social maturity index. The results highlighted that six of the eight IPB scales demonstrated positive significant correlations with the CPI–social maturity index, with only the autonomy and intimacy scales exhibiting non-significant coefficients (Hannah et al., 1996).

MOS 36-Item Short-Form Health Survey (SF-36)

The SF-36 is a 36-item scale to assess the health-related QOL, with higher scores indicating better QOL. The SF-36 involved eight subscales: physical functioning (PF), role physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role emotional (RE), and mental health (MH). Through applying special algorithms, these eight subscales could contribute to a physical dimension called the physical component scale (PCS), and a mental dimension called the mental component scale (MCS) (Sherbourne and Ware, 1992).

The SF-36 Taiwan version was a popular scale in which internal reliability has reached acceptable level for all scales ($\alpha > 0.70$). The Cronbach's α of eight subscales ranged from 0.65–0.92 (PF = 0.91, RP = 0.92, BP = 0.76, GH = 0.82, VT = 0.79, SF = 0.65, RE = 0.87, and MH = 0.78). In terms of validation, each item showed item-scale correlation higher than 0.4. Except item 5 in MH, every item had higher correlation coefficient with its own subscale than other subscales. This indicated good discriminant validity. SF-36 has Taiwanese version and is used commonly in Taiwan for years. The item-scale correlation coefficients of SF-36 range from 0.40 to 0.83 and the rest of the scales have passed the tests of item discriminant validity, except for MH5 (Lu et al., 2003; Tseng et al., 2003).

METHODOLOGY

This study was approved by an institutional review board of China Medical University in Taiwan (Approval No. DMR95-IRB-144). There are various methods for crosscultural adaptation of a questionnaire, but none is considered as the gold standard (Epstein et al., 2015). This study integrated the research methods of Beaton et al. (2000) and Rived-Ocaña et al. (2020) and the steps are shown in **Figure 1**.

The first step was to develop the 120 items of the Chinese IPB (C-IPB). The process of translation and back-translation was used, and expert validity and reliability test in a pilot study were done.

The second step was to construct C-IPB short-form (CIPB-SF). This step involved item analysis and factor analysis. The item analysis was used to delete those items without discrimination and with poor homogeneity. The left items were tested in factor analysis, and items which showed factor loading more than 0.55 remained, and the initial CIPB-SF was formed.

Finally, the reliability and validity of the CIPB-SF were examined. The reliability was tested in internal consistency, and the convergent validity, confirmatory factor analysis (CFA), and criterion-related validity were done.

The First Step: Establish the Chinese Inventory of Psychosocial Balance

After obtaining permission from the author of the original IPB questionnaire, the questionnaire was translated following the process recommended by Beaton et al. (2000) and Rived-Ocaña et al. (2020). The first step included "translation, back-translation," "expert validity," and "reliability of pilot study."

Translation, Back-Translation

The first translator was a professional translator with a master's degree and experience in English-to-Chinese translation. After the English-to-Chinese translation, another professional translator back-translated the first Chinese version of the IPB into English. Both translators were Chinese, and they did not engage in any discussions with each other. Thus, both professional translators produced translations of the IPB independently.

Expert Validity

Subsequently, expert validity was done. The expert was an academic scholars or a clinical staff with a master or doctoral degree and specialized in psychological or clinical care. We invited five experts (a doctoral degree of nursing, a master degree of psychology, a doctoral degree of social work, a master degree of nursing, and master degree of nursing supervisor) who reviewed both the versions and offered suggestions regarding the questionnaire content, semantics, structure, syntax, orthographic revision, and appropriateness of the translation. They were anonymized to each other during the process. The content validity index (CVI) was used in assessing the content validity. The CVI was a widely useful approach for content validity in instrument development and it could be computed *via* the item-CVI (I-CVI) and the scale-level-CVI (S-CVI). The experts were asked to rate instrument items on a 4-point ordinal scale [4(highly relevant), 3(quite relevant), 2(not relevant), 1(extreme not relevant)]. Item-CVI (I-CVI) is computed as the number of experts who score a rating of 3 or 4 point of each item, divided by the total number of experts. The scale-level CVI (S-CVI) is defined as "the proportion of items on an instrument that achieved a rating of 3 or 4 by the content experts" (Beck and Gable, 2001). After integrating the opinions from five experts, the corrections were done and the revised IPB was returned to the experts for second review. Until every expert agreed the contents of translated IPB, the 120-item C-IPB was completed.

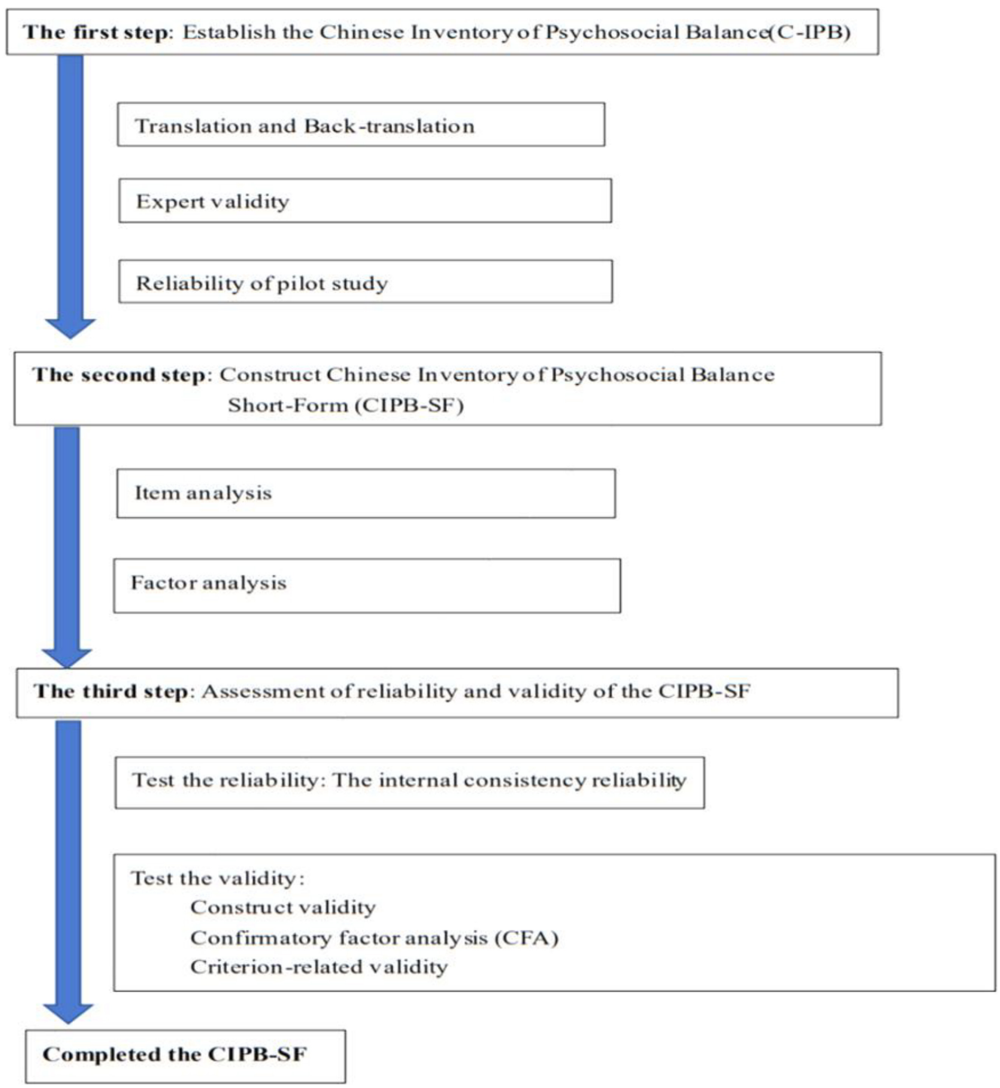

FIGURE 1 | Development and establishing the CIPB-SF.

Reliability of Pilot Study

Following the expert validity, a pretest was conducted to identify the problems that could be encountered while using the scale in the study and to use the responses obtained during the pretest for consultation purposes in the future. During the pretest, a majority of the participants expressed that filling out the 120-item questionnaire took them 30–45 min and that the long questionnaire tired them. To facilitate economic efficiency and widespread usage, we developed a short form of the C-IPB, called the CIPB-SF.

The Second Step: Constructing Chinese Inventory of Psychosocial Balance Short-Form

In the second step, the CIPB-SF was developed *via* item analysis and factor analysis.

Item Analysis

Item discrimination and a test of homogeneity were used to perform the item analysis. Theoretically, items in a scale should have the capability to distinguish good from bad condition of subjects and those items in the same stage had the same characteristics. Therefore, according to the score on C-IPB, we classified the top 27% of the sample as the high group and the bottom 27% as the low group, and *t*-test was conducted to test if there were differences between the high group and the low group. The items which were not statistically significant implied that the items lacked discrimination and had to be deleted. The homogeneity test was analyzed using correlation coefficients. Correlation coefficients with magnitudes less than 0.3 had little correlation with other items in the same stage. Items were deleted when the correlation coefficients were less than 0.3 (Williams et al., 2010). Based on the above item analysis, the items with discriminative and homogeneous of the scale were entitled in the factor analysis.

Factor Analysis

Before factor analysis, we performed Bartlett's test of sphericity and the Kaiser–Meyer–Olkin measure of sampling adequacy test (KMO test). The method of principal component analysis and varimax rotation were used. The criterion of eigenvalues > 1 was employed to select components (Munro, 2005). Items with factor loadings that exceeded 0.55 were included (Williams et al., 2010; Demoulin, 2016). Finally, the CIPB-SF was established.

The Third Step: Assessment of Reliability and Validity of the Chinese Inventory of Psychosocial Balance Short-Form

In the third step, we assessed the reliability and validity of the CIPB-SF. We calculated the Cronbach's α to measure the internal consistency reliability and compare the Cronbach's α between the CIPB-SF and the C-IPB. When the Cronbach's α scores were > 0.70, the reliability was considered to be high. The Cronbach's α scores between 0.35 and 0.7 indicated acceptable reliability (Taber, 2018).

Furthermore, we tested the construct validity, CFA, and criterion-related validity of CIPB-SF.

"Convergent validity" is one of the construct validity and which states that the items having the similar constructs should be highly correlated. One of the method is calculating the correlation coefficients between tools' subdomains that are considered to measure the same construct (Gregory, 2004).

We tested the "CFA" of CIPB-SF *via* structural equation modeling (SEM). Based on the literature, it is suggested that if the sample size over 500 would be more adequate to use the maximum likelihood (ML) method for estimation (Long, 1997). Due to the sample size in the research was only 308 and the questionnaires were scored as continuous variables, the generalized least square (GLS) method was conducted in the study (Carroll and Ruppert, 1982).

In order to evaluate the "criterion-related validity" of CIPB-SF, we planned to compare the CIPB-SF and SF-36 *via* Pearson's correlation. The original scale, IPB, was compared with the social maturity index for testing the criterion-related validity of IPB (Domino and Affonso, 1990). Though, the Chinese-form social maturity index does not exist, we still have to evaluate the CIPB-SF with the similar concept and Chinese-form scale. Theoretically, participants with better ego development will have higher scores in QOL (Kleijn et al., 2016). Therefore, we compared CIPB-SF and QOL through Pearson correlation to analyze the criterion-related validity of CIPB-SF.

RESULTS

General Information of Participants

Since the Erikson's theory was applied for the research, the elder adults were qualified for all the eight stages of ego development only. Residential older adults older than 65 years old were contacted by a student pursuing a master's degree in nursing. Since our study concerned general older adults, we wished to recruit healthy Taiwanese older adults who were also independent. Therefore, the data collection sites were settled to public places. Individuals with dementia or other geriatric cognitive disorders and those who lived in long-term care facilities were not eligible.

Participants were approached by a researcher who explained the study purpose and content, and they signed informed consents before data collection. The mini-mental state examination (MMSE) was developed in 1975 by Folstein and Mc Huge, and the evaluating dimensions included orientation to time and place, registration, attention and calculation, recall, language, repetition, and complex commands. The total score is 30, and a higher value of MMSE indicates a better recognition. MMSE score less than 23 means mild cognition impairment. MMSE is the most used measurement tool in cognitive functions in clinical settings (Galea and Woodward, 2005; Finney et al., 2016). This study applied the MMES as the screening tool, and MMSE scores under 23 are represented as the cut-off score for cognitive impairment. Those participants who had MMSE scores below 23 were excluded in the data analysis (Galea and Woodward, 2005).

A total of 308 questionnaires were retrieved. Participants' mean age was 71 years (standard deviation = 6.4). In addition, 52.3% of them were women and 96.4% were married. Regarding education levels, 18.5% of participants were uneducated, 37.3% have received high school education or above. In terms of marriage, only 11 participants were never married. Most participants were married or divorced or widowed. There were 52.9% participants who were religious, and 63% participants felt their economic statuses were ordinary. In **Table 1**, the participants' socio-demographic characteristics are presented.

Establish the Chinese Inventory of Psychosocial Balance

After translation and back-translation, the translated Chinese version IPB and English version IPB were dispatched to five experts for check and reviewing. The content, semantics, structure, syntax, orthographic revision, and appropriateness of the translated IPB were carefully examined and considered with revision suggestions. After the first-time revision, the revised Chinese IPB was returned to the experts for further review. When consensus was reached, the CVI was calculated.

The CVI for the C-IPB was 0.7, indicating the scale's acceptable expert validity (Polit and Beck, 2004). After conducting a final review for content validity, all experts agreed with the scale's items. Thus, the 120-item scale, C-IPB, was completed.

A pretest was conducted to identify the problems that could be encountered while using the scale in the study and to use the responses obtained during the pretest for consultation purposes in the future. We included 32 subjects who were ≥ 65 years and without any cognitive impairment in the pretest. The pretest results revealed the following: (1) The reliability of the C-IPB was good (Cronbach's α = 0.82). (2) The length of the 120-item C-IPB was too long for the elderly participants to complete. Many participants mentioned that completing the 120-item C-IPB tired

TABLE 1 | Participant socio-demographic characteristics (*N* = 308).

		Mean	*SD*	n	Percentage
Age		71	6.4		
Gender					
	Male			147	47.7
	Female			161	52.3
Education					
	Uneducated			57	18.5
	Elementary and junior			136	44.2
	High school graduate or above			115	37.3
Marriage					
	Yes			297	96.4
	Never			11	3.6
Religious					
	Yes			163	52.9
	No			145	47.1
Self-perceived economic situation					
	Poor			37	12.0
	General			194	63.0
	Rich			25	8.1
	Rejection			52	16.9

them. Therefore, we decided to shorten the C-IPB on the basis of the pretest results to increase its usefulness.

The C-IPB possessed good reliability with Cronbach's α = 0.82 and face validity CVI = 0.7; in the meanwhile, we found that completing 120 items is a burden to participants; thus simplifying the C-IPB was necessary.

Construct Chinese Inventory of Psychosocial Balance Short-Form

Item Analysis of the Chinese Inventory of Psychosocial Balance

A method for the comparison of extreme groups was tested. Items 10, 12, 17, 49, 74, 76, 83, 87, and 117 were not statistically significant, implying that those items lacked discrimination and had to be deleted. The remaining 109 items possessed discrimination.

The homogeneity test was analyzed using correlation coefficients. On the basis of the homogeneity test results, 74 items were deleted from the following stages because of having correlation coefficients below 0.3: trust stage (eight items deleted), autonomy stage (10 items deleted), initiative stage (nine items deleted), industry stage (nine items deleted), identity stage (10 items deleted), intimacy stage (nine items deleted), generativity stage (10 items deleted), and integrity stage (nine items deleted).

The result indicated that items 10, 12, 17, 49, 74, 76, 83, 87, and 117 lacked both discrimination and homogeneity and had to be deleted. A total of 46 items were reserved through item analysis and were included in factor analysis.

Factor Analysis of the Chinese Inventory of Psychosocial Balance

Before factor analysis, we performed the KMO test. The KMO test results for the C-IPB was 0.86, which indicated that sampling is adequate (Kaiser and Rice, 1974). Through principal component analysis and varimax rotation, eight factors could be extracted from the ego development. Items with factor loadings that exceeded 0.55 were included (Williams et al., 2010; Demoulin, 2016). **Table 2** presents the factor analysis results. Finally, we retained five items of each stage in order to avoid confusion for the user in the future.

Assessment of Reliability and Validity of the Chinese Inventory of Psychosocial Balance Short-Form

Reliability Coefficients for the Chinese Inventory of Psychosocial Balance Short-Form

Table 3 presents the results. The reliability of the CIPB-SF was found to be adequate, with internal consistency and a Cronbach's α = 0.85. Compared to the CIPB-SF and the C-IPB, the Cronbach's α of each stage was similar. The Cronbach's α of each stage ranged from 0.81 to 0.89 of the CIPB-SF, and from 0.70 to 0.77 of the C-IPB. **Table 3** showed the total score and subscores of the eight stages. All stages exhibited sufficient internal consistency, as indicated by good Cronbach's α coefficients. The highest Cronbach's α coefficient was 0.89 in the identity stage, and the lowest was 0.81 in the trust and integrity stages.

Criterion-Related Validity of Chinese Inventory of Psychosocial Balance Short-Form

We evaluated the Pearson's correlation coefficient between the CIPB-SF and SF-36 (**Table 4**). The CIPB-SF subscales all had a significant positive relation ($p < 0.001$) to the MCS of SF-36. The results indicated that the better ego development status the individuals have, the better mental dimension of QOL they will have. The result was in conformity with the hypothesis (Kleijn et al., 2016), and it means the CIPB-SF had a good criterion-related validity.

Validity of the Chinese Inventory of Psychosocial Balance Short-Form

Table 4 presents the correlations among each item of the subscale and other stages in the CIPB-SF. All subscales of the CIPB-SF had significant correlations with their own stage, and most of the correlation coefficients were > 0.4 except, item 67 (from the initiative stage), items 69 and 101 (from the identity stage), and items 46, 86, and 6 (from the intimacy stage). The three stages of initiative, intimacy, and identity had fair validity.

The eight subscales of the CIPB-SF and their total scores exhibited significant moderate to high positive correlations with their own stage. All of the items had significant positive correlations with their own stage, with greater correlation coefficient than others (**Table 4**).

TABLE 2 | Factor analysis of C-IPB (*N* = 308).

Component	Factor loading							
	Factor1	**Factor2**	**Factor3**	**Factor4**	**Factor5**	**Factor6**	**Factor7**	**Factor8**
"Trust vs. Mistrust"								
9. I have confidence in my own abilities	**0.73**	0.63	0.00	0.12	0.23	0.32	0.05	0.00
41. Most conflicts between people can be resolved by discussion	**0.72**	0.67	0.05	0.04	0.01	0.09	0.03	0.17
113. Basically, I think I am an all right person	**0.64**	0.50	0.02	0.23	0.27	0.27	0.04	0.11
97. In general I am optimistic person	**0.58**	0.21	0.27	0.13	0.21	0.21	0.18	0.05
33. Suffering can be meaningful for the growth of person	**0.57**	0.10	0.32	0.27	0.02	0.20	0.04	0.01
105. People have the capacity to solve their problems	**0.57 (delete)**	0.48	0.04	0.14	0.08	0.10	0.17	0.33
*89. I find I am open to new ideas	**0.54 (delete)**	0.48	0.21	0.01	0.22	0.02	0.11	0.03
"Autonomy vs. Shame and doubt"								
114."A place for everything and everything in its place" is my motto	0.15	**0.78**	0.54	0.03	0.05	0.10	0.10	0.15
34. I am a very organized person	0.15	**0.71**	0.02	0.08	0.20	0.56	0.26	0.12
82. We would all be better off if people obeyed the laws we have	0.16	**0.70**	0.30	0.14	0.00	0.58	0.00	0.03
66. It is important for young people to be independent	0.02	**0.59**	0.09	0.47	0.23	0.38	0.13	0.00
58. I am quite self-sufficient	0.51	**0.58**	0.26	0.25	0.10	0.20	0.09	0.14
"Initiative vs. Guilt"								
107. It's easy for me to begin new projects	0.18	0.04	**0.71**	0.54	0.14	0.26	0.18	0.01
115. I am a highly curious individual	0.05	0.07	**0.65**	0.05	0.46	0.33	0.13	0.12
51. As a child I often took things apart to see how they worked	0.19	0.02	**0.65**	0.34	0.23	0.30	0.27	0.10
67. I would love to invent a new way of doing something	0.02	0.09	**0.61**	0.23	0.41	0.23	0.05	0.10
99. Friends and acquaintances often tell me that my ideas are original	0.29	0.11	**0.60**	0.25	0.24	0.24	0.22	0.10
3. I am easily embarrassed	0.10	0.11	**0.59 (delete)**	0.16	0.48	0.08	0.20	0.24
"Industry vs. Inferiority"								
116. I genuinely enjoy work	0.49	0.13	0.07	**0.72**	0.02	0.16	0.01	0.09
108. I like to be busy	0.01	0.34	0.20	**0.71**	0.09	0.46	0.01	0.10
92. When necessary, I can devote a lot of energy to a task	0.21	0.11	0.10	**0.70**	0.45	0.07	0.15	0.14
28. Others would describe me as a productive person	0.49	0.01	0.09	**0.67**	0.02	0.16	0.13	0.07
52. Work brings me great satisfaction	0.05	0.19	0.04	**0.67**	0.45	0.01	0.04	0.04
4. Most people around me seem to be more talented than I am	0.35	0.23	0.01	**0.62 (delete)**	0.14	0.27	0.52	0.03
"Identity vs. Role confusion"								
69. I feel very comfortable with the value I have	0.21	0.33	0.14	0.02	**0.75**	0.14	0.04	0.02
5. Sometimes I wonder who I really am	0.05	0.39	0.07	0.16	**0.65**	0.07	0.07	0.09
109. Friends would describe me as a very changeable person	0.52	0.27	0.02	0.15	**0.62**	0.20	0.12	0.10
93. My adolescence was fairly stormy	0.20	0.14	0.14	0.09	**0.58**	0.13	0.37	0.35
101. There are times when I wish I had been born of the opposite sex	0.12	0.26	0.21	0.11	**0.55**	0.15	0.27	0.02
"Intimacy vs. Isolation"								
46. There have been people in my life with whom I have been willing to share my innermost thoughts	0.04	0.12	0.23	0.12	0.09	**0.66**	0.49	0.28
86. Overall, my sexual life has been satisfactory	0.41	0.32	0.08	0.05	0.12	**0.66**	0.47	0.14
102. When I was a teenager I had a very close friend with whom I shared many experiences	0.13	0.09	0.41	0.36	0.14	**0.66**	0.30	0.27

(Continued)

TABLE 2 | (Continued)

Component	Factor loading							
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8
38. There have been times when I felt extremely close to someone I loved	0.26	0.00	0.39	0.15	0.06	**0.60**	0.20	0.28
6. I have experienced some very close friendships	0.16	0.29	0.11	0.33	0.20	**0.58**	0.18	0.41
110. I enjoy being with people	0.05	0.26	0.17	0.17	0.02	**0.55 (delete)**	0.54	0.03
"Generativity vs. Stagnation"								
119. I enjoy learning new skills	0.06	0.06	0.07	0.27	0.11	0.09	**0.79**	0.42
47. To be a good parent is one of the most challenging tasks people face	0.13	0.07	0.24	0.16	0.17	0.18	**0.68**	0.52
39. If it were possible, I would greatly enjoy teaching adolescents	0.15	0.08	0.27	0.17	0.25	0.09	**0.62**	0.53
63. I am very concerned that our children will grow in a polluted world	0.15	0.07	0.17	0.16	0.06	0.09	**0.60**	0.42
7. I derive great pleasure in watching a child master a new skill	0.13	0.10	0.29	0.15	0.18	0.13	**0.57**	0.53
"Integrity vs. Despair"								
56. There are many things I enjoy in life	0.53	0.08	0.21	0.08	0.10	0.29	0.10	**0.75**
40. Life has been good to me	0.01	0.49	0.11	0.21	0.35	0.08	0.02	**0.74**
48. I have left my mark on the world	0.47	0.15	0.36	0.20	0.07	0.09	0.05	**0.71**
64. I find little sense in living	0.40	0.48	0.05	0.07	0.16	0.05	0.17	**0.59**
96. When I die I will be missed	0.47	0.12	0.06	0.09	0.06	0.23	0.22	**0.57**
104. I have given serious thought to the meaning of life	0.16	0.20	0.10	0.10	0.18	0.53	0.06	**0.56 (delete)**

The bold values represented the factor loading values in the same group for better readibility.

TABLE 3 | Reliability coefficients (Cronbach's α) for the CIPB-SF and C-IPB (N = 308).

	CIPB-SF	C-IPB
Stage_1	0.81	0.70
Stage_2	0.82	0.71
Stage_3	0.84	0.77
Stage_4	0.82	0.76
Stage_5	0.89	0.77
Stage_6	0.85	0.76
Stage_7	0.82	0.70
Stage_8	0.81	0.72
Eight stages	0.85	0.74

Stage_1: The first stage (trust vs. mistrust); Stage_2: The second stage (autonomy vs. shame and doubt).
Stage_3: The third stage (initiative vs. guilt); Stage_4: The fourth stage (industry vs. inferiority).
Stage_5: The fifth stage (identity vs. role confusion); Stage_6: The sixth stage (intimacy vs. isolation).
Stage_7: The seventh stage (generativity vs. stagnation); Stage_8: The eighth stage (integrity vs. despair).
Eight stage: A composite score of the eight subscales.

Confirmatory Factor Analysis of the Chinese Inventory of Psychosocial Balance Short-Form

The aim of the CFA was to verify the models of questionnaires by structural equation (Olmedo Moreno et al., 2014). Based on the results of item analysis and factor analysis of the C-IPB, forty items were retained as the CIPB-SF and we defined those as Model 1 (**Figure 2**). The CIPB-SF we developed showed good reliability in all eight stages and appropriate validity, except the fifth and sixth stages. Therefore, the short-form fifth and sixth stages scale were replaced by the original fifth and sixth stages scale of IPB, and we defined those as Model 2 (**Figure 3**). Both the models will be evaluated by model fit index for assessment. As shown in **Table 5**, the measurement model showed an acceptable data fit included χ^2/degrees of freedom, root mean square error of approximation (RMSEA), and Parsimony goodness-of-fit index (PGFI). However, the root mean square residual (RMR) was above the critical limit, 0.05, and the GFI was slightly less than the standard level. Sharma et al. (2005) indicated that the GFI was affected by sample size. The larger sample size should be considered in the future researches.

Akaike information criterion is used for model selection, and the model with lower value means that is better fit for the data. In the study, Model 1 is a better fit for the data compared to the Model 2 (**Table 5**).

DISCUSSION

In this study, we translated the IPB into Chinese to construct the CIPB-SF through factor analysis. The results obtained confirmed that the SC-IPB is a valid and reliable comprehensive tool that can be used to evaluate ego development among the Chinese elderly population. The previous studies indicated that the length of a questionnaire is important because it can directly affect response rates, survey costs, and data quality (Lavrakas, 2008

TABLE 4 | Correlation among each item of subscale, all stages in CIPB-SF, and MCS (N = 308).

Stage	MCS	Eight stage	Stage_1	Stage_2	Stage_3	Stage_4	Stage_5	Stage_6	Stage_7	Stage_8
Item	**R**	**r**	**r**	**r**	**r**	**r**	**r**	**r**	**r**	**r**
Stage_1	0.47**									
9		**0.62**	0.72	0.51	0.50	0.62	0.04	0.29	0.43	0.49
41		**0.50**	0.56	0.54	0.24	0.46	0.06	0.26	0.51	0.33
113		**0.64**	0.69	0.57	0.36	0.50	0.15	0.40	0.50	0.54
97		**0.46**	0.57	0.31	0.43	0.34	−0.06	0.39	0.43	0.28
33		**0.53**	0.59	0.36	0.36	0.49	0.19	0.27	0.26	0.52
Stage_2	0.43**									
114		**0.51**	0.50	0.69	0.33	0.51	0.02	0.21	0.42	0.39
34		**0.52**	0.47	0.63	0.26	0.45	0.29	0.16	0.29	0.51
82		**0.43**	0.41	0.61	0.15	0.39	0.24	0.11	0.31	0.36
66		**0.53**	0.46	0.69	0.29	0.40	0.11	0.33	0.50	0.36
58		**0.60**	0.54	0.75	0.43	0.54	0.09	0.28	0.51	0.44
Stage_3	0.24**									
107		**0.46**	0.36	0.36	0.61	0.42	−0.07	0.33	0.39	0.32
115		**0.41**	0.29	0.27	0.67	0.20	−0.01	0.42	0.35	0.25
51		**0.41**	0.35	0.25	0.60	0.37	−0.07	0.31	0.40	0.25
67		0.28	0.23	0.11	0.60	0.20	−0.14	0.24	0.25	0.23
99		**0.56**	0.49	0.36	0.69	0.46	−0.04	0.43	0.49	0.40
Stage_4	0.45**									
116		**0.49**	0.45	0.47	0.39	0.63	0.01	0.27	0.40	0.29
108		**0.58**	0.55	0.54	0.30	0.69	0.18	0.25	0.41	0.47
92		**0.50**	0.44	0.43	0.33	0.67	−0.04	0.28	0.40	0.44
28		**0.59**	0.62	0.58	0.32	0.67	0.08	0.30	0.39	0.49
52		**0.45**	0.43	0.32	0.34	0.72	0.00	0.19	0.36	0.29
Stage_5	0.45**									
69		0.04	0.01	0.09	−0.21	0.01	0.65	−0.20	−0.14	0.20
5		**0.46**	0.18	0.20	−0.01	0.16	0.71	0.01	0.01	0.32
109		**0.48**	0.11	0.13	0.06	0.02	0.61	−0.03	0.02	0.25
93		**0.47**	0.04	0.15	0.03	0.06	0.57	0.03	0.09	0.18
101		0.04	−0.03	0.13	−0.22	0.03	0.60	−0.17	−0.07	0.10
Stage_6	0.24**									
46		0.22	0.13	0.03	0.34	0.10	−0.20	0.64	0.31	0.05
86		0.24	0.17	0.03	0.22	0.04	−0.07	0.64	0.25	0.18
102		**0.47**	0.42	0.38	0.38	0.32	−0.11	0.60	0.44	0.35
38		**0.51**	0.38	0.25	0.39	0.35	0.05	0.63	0.41	0.48
6		0.31	0.21	0.13	0.39	0.14	−0.13	0.67	0.28	0.20
Stage_7	0.31**									
119		**0.52**	0.47	0.45	0.42	0.44	−0.03	0.40	0.63	0.31
47		**0.47**	0.42	0.39	0.38	0.34	0.05	0.35	0.57	0.33
39		**0.49**	0.42	0.33	0.36	0.42	−0.04	0.35	0.64	0.45
63		**0.41**	0.30	0.37	0.29	0.28	−0.01	0.33	0.66	0.32
7		**0.61**	0.57	0.42	0.55	0.49	−0.05	0.47	0.75	0.39
Stage_8	0.53**									
56		**0.58**	0.58	0.45	0.31	0.42	0.26	0.24	0.42	0.70
40		**0.50**	0.40	0.34	0.33	0.34	0.24	0.25	0.37	0.66
48		**0.65**	0.52	0.54	0.43	0.49	0.20	0.41	0.46	0.73
64		**0.41**	0.27	0.38	0.10	0.34	0.45	0.08	0.21	0.63
96		**0.48**	0.48	0.29	0.30	0.38	0.17	0.32	0.26	0.58

$^{**}p<0.01$.
Stage_1: The first stage (trust vs. mistrust); Stage_2: The second stage (autonomy vs. shame and doubt).
Stage_3: The third stage (initiative vs. guilt); Stage_4: The fourth stage (industry vs. inferiority).
Stage_5: The fifth stage (identity vs. role confusion); Stage_6: The sixth stage (intimacy vs. isolation).
Stage_7: The seventh stage (generativity vs. stagnation); Stage_8: The eighth stage (integrity vs. despair).
Eight stage: A composite score of the eight subscales. MCS, Mental component scale of SF-36.
The bold represented the values which were larger than 0.4.

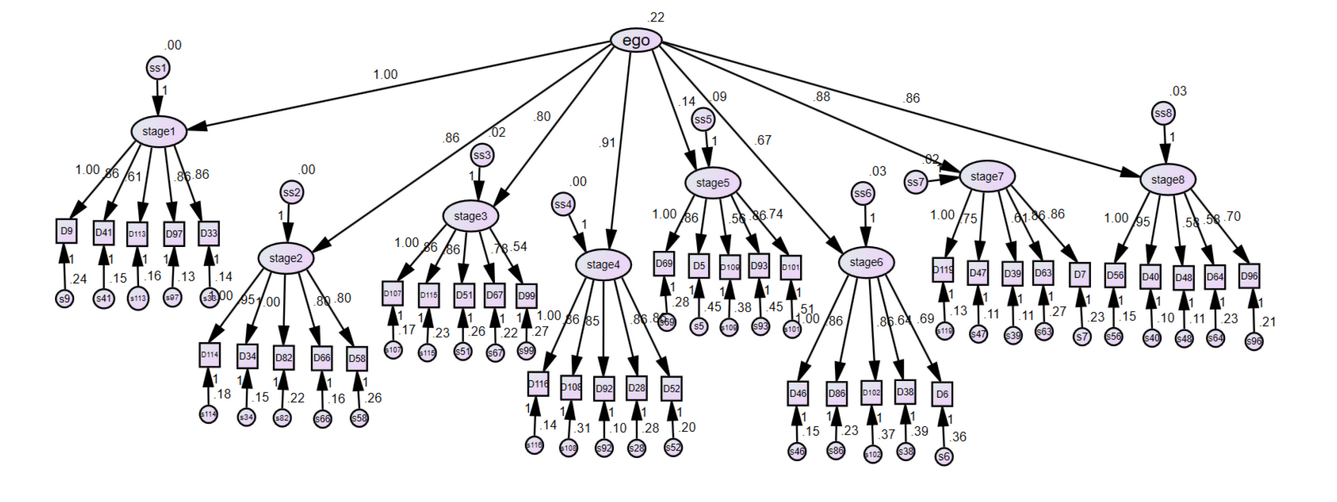

FIGURE 2 | The model 1.

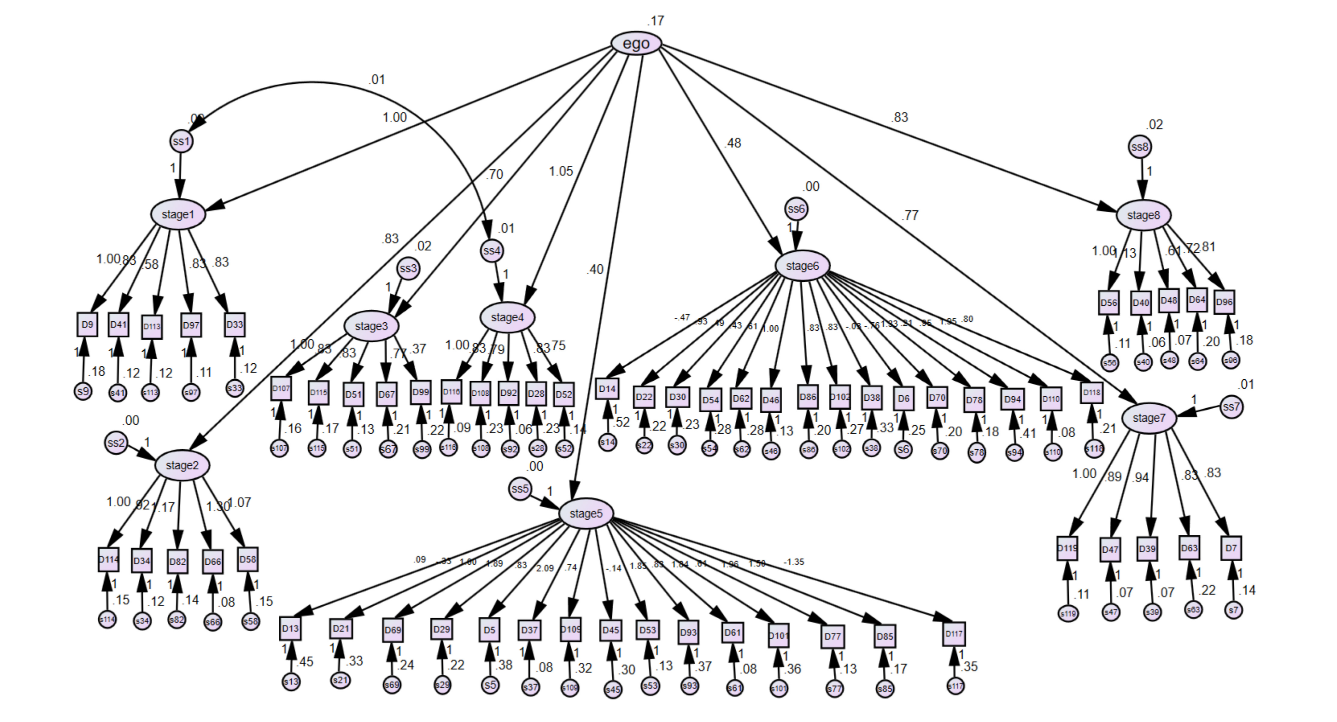

FIGURE 3 | The model 2.

TABLE 5 | Comparative summary of model fit.

	χ^2	df	χ^2/df	RMR	GFI	PGFI	CFI	TLI	RMSEA	AIC
Model 1	1501.21	746	2.01	0.09	0.76	0.69	0.82	0.53	0.05	1649.21
Model 2	3099.611	1715	1.81	0.10	0.67	0.62	0.62	0.61	0.05	3329.61

"Encyclopedia of Survey Research Methods,"). We shortened the 120-item IPB to a 40-item CIPB-SF with similar reliability and validity. In its ability to avoid any unnecessary interference from the questionnaire completion being a time-consuming task, the 40-item CIPB-SF can be considered superior to the original scale.

In this study, the CIPB-SF was found to have good reliability. Moreover, the Cronbach's α of each of the eight subscales was between 0.81 and 0.89. We compared the reliability of the CIPB-SF with that of the IPB. The Cronbach's α in the CIPB-SF was greater than that in the IPB. The values of the Cronbach's α coefficients in the IPB were between 0.64 and 0.79 (Domino and Affonso, 1990). In addition, the highest Cronbach's α coefficient value was observed in the industry stage and the lowest was observed in the intimacy stage of the IPB. In this study, the highest value of Cronbach's α coefficients of the CIPB-SF was observed in the identity stage. Although the trust and integrity stages had the lowest Cronbach's α coefficient values of the CIPB-SF, the Cronbach's α coefficients of the eight stages of the CIPB-SF had similar values, ranging between 0.81 and 0.89. The autonomy stage (Cronbach's α = 0.65) had the lowest Cronbach's α value in the IPB.

The 30–35-day test–retest reliability in the subscale scores of the IPB were < 0.50 (Domino and Affonso, 1990). During the CIPB-SF development process, the subjects who participated in the study were anonymously recruited from the community. Therefore, assessing the test–retest reliability was not possible.

The validity of the IPB was assessed by examining the associations of the IPB subscales to the CPI–social maturity index (Hannah et al., 1996). Six of the eight IPB scales were found to have a positive significant correlation with the CPI-social maturity index, with only the autonomy and the intimacy stage exhibiting coefficients (Hannah et al., 1996). The CIPB–SF of the study had acceptable validity, except for the intimacy and identity stages. The validity testing results of the CIPB-SF were similar to those of the IPB. The reliability and validity of the identity stage differed between the two questionnaires, IPB and CIPB-SF. Cultural differences between the Eastern and Western versions were expected. The older adults who were a product of the Taiwanese experiment in Progressionism must have studied hard during their adolescence period to succeed in examinations for entering higher education. During their youth, these older adults had also witnessed the colonial rule of Japan. Therefore, they must have faced racial discrimination and unequal treatment in society. The "monomania to study medicine" and strong parental authority also emerged in their era (Chen, 2010). Cultural differences may have resulted in negative consequences, especially in the identity stage.

The intimacy vs. isolation stage had the lowest quality in the CIPB-SF in terms of convergent and discriminant validity. The CIPB-SF was developed on the basis of the IPB; however, the social backgrounds of these scales pertain to differ in terms of the expression of love. According to a study that explored Taiwanese male older adults through narrative inquiry, children born in Taiwan in the 1940s tend to have an initial life attitude of fatalism, but are not willing to yield to fate. This finding also highlights that men attach considerable importance to personal achievements and career development (Chuang and Lin, 2017). In view of the aforementioned culture differences, we recommend future researchers to perform another advanced study on the intimacy vs. isolation stage using qualitative research methods.

In CFA, the GFI values were 0.76 in Model 1 and 0.67 in Model 2. This result indicated that the model fit is only fair but not in good-fit. This may due to the translation process, under crossculture context, and reduction of the number in items. For instance, in Eastern, people are more conserved in showing intimacy and love than in Western culture. These factors may reduce the grade of model fit. This CIPB-SF could be tested and modified in future studies for developing a more suitable measurement tool based on Erikson theory.

To construct the IPB, 528 subjects, including students, adults, elderly individuals, and retired adults, between the ages of 21 and 88 years were included in the study (Domino and Affonso, 1990). A total of 308 subjects aged ≥ 65 years who participated in the development of the CIPB-SF were different from those recruited for the IPB. The IPB is a well-tested questionnaire, and the CIPB-SF was developed following the original questionnaire, which was first translated and then shortened. On the basis of Erikson's theory and the IPB, scales pertaining to individual stages may also be used in isolation. Scales concerning individual stages that are tested in corresponding groups should be considered in the future.

Addressing the weaknesses of the present study, the first is the present research that used a cross-sectional design. Undoubtedly, it would be desirable to acknowledge the present study *via* applying a longitudinal study design. However, the IPB is already a well-tested questionnaire, and the present study focused emphatically on contributing to the Chinese-form questionnaire.

Then, the present study excluded the older adults who lived in long-term care institution. There was some discrepancy among older adults who lived in long-term care institution and community. Üzar-Özçetin and Ercan-Şahin (2020) indicated that older adults who have to live in nursing homes may feel deprived and isolated from society. To avoid the interference of the sampling, we recruited healthy Taiwanese elderly adults who were also independent. However, further research is needed to determine institutional residents' validity or crosscultural validity.

The CIPB-SF can be used as a tool to investigate the status of ego development in the elderly community. Finally, future studies should implement and test CIPB-SF among different subjects that enhance the applicability and practicality of CIPB-SF.

CONCLUSION

Erikson's ego development theory is a rare psychosocial theory that encompasses the entire life span ranging from infancy to late adulthood. However, the lack of a credible Chinese version of the scale may impede research on ego development in Taiwan. This study constructed a 40-item CIPB-SF with good reliability (Cronbach's α = 0.81–0.89) and validity. Due to its conciseness, the 40-item CIPB-SF was more appropriate to apply in for the Chinese elderly population to avoid physical overload. This scale can also serve as a useful tool for convenient screening in the future.

LIMITATION

There was some limitation in this research. First, the sample size of this study was less than 500 subjects, yet the internal consistency test of CIPB-SF and the eight stages were still good (Cronbach's α = 0.81–0.89). We will invite other research teams to participate in the study related to CIPB-SF in the future. Second, the low levels of CFA comparative indices (CFI, GFI, and TLI), although an absolute fit index like RMSEA were good. Although the comparative indices (CFI, GFI, TLI) indicated that the model fit was fair, Sharma et al. (2005) pointed out that the GFI was affected by sample size, and our sample size was less than 500, which is not perfect in SEM analysis. Further study will be needed to recruit multiple research teams to cooperate in crossdomain and crosscultural validity.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board (IRB) of China Medical University Hospital in Taiwan approved the study. The committee's reference number is "DMR95-IRB-144." Written informed consent was obtained from participants. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

P-YC wrote the main manuscript text, performed the research, and analyzed the data. CL interpreted the data. W-CH checked the statistical process. T-PY revised the sentence and checked the grammar. All authors reviewed the manuscript.

FUNDING

This research was funded by the China Medical University Hospital (Grant No. DMR-106-147).

REFERENCES

Ardelt, M., Gerlach, K. R., and Vaillant, G. E. (2018). Early and midlife predictors of wisdom and subjective well-being in old age. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 73, 1514–1525. doi: 10.1093/geronb/gby017

Beaton, D. E., Bombardier, C., Guillemin, F., and Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25, 3186–3191. doi: 10.1097/00007632-200012150-00014

Beaumont, S. L., and Pratt, M. M. (2011). Identity processing styles and psychosocial balance during early and middle adulthood: the role of identity in intimacy and generativity. *J. Adult Dev.* 18, 172–183. doi: 10.1007/s10804-011-9125-z

Beck, C. T., and Gable, R. K. (2001). Ensuring content validity: an illustration of the process. *J. Nurs. Meas.* 9, 201–215. doi: 10.1891/1061-3749.9.2.201

Brennan, M., and MacMillan, T. (2006). Developmental recapitulation in adaptation to vision loss among middle-age and older adults. *J. Soc. Work Disabil. Rehabil.* 5, 45–63. doi: 10.1300/J198v05n01_03

Busch, H., Hofer, J., Poláčková Šolcová, I., and Tavel, P. (2018). Generativity affects fear of death through ego integrity in German, Czech, and Cameroonian older adults. *Arch. Gerontol. Geriatr.* 77, 89–95. doi: 10.1016/j.archger.2018.04.001

Carroll, R. J., and Ruppert, D. (1982). A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model. *J. Am. Stat. Assoc.* 77, 878–882. doi: 10.1080/01621459.1982.10477901

Chen, C.-K. (2010). The social meaning and influence of the Taiwanese "monomania to study medicine" in the Japanese colonial period. *Cult. Pract. Soc. Change* 1, 127–195. doi: 10.30125/c.201006.0004

Chuang, Y.-H., and Lin, C.-S. (2017). Narrating life images of a self-made senior citizen through life review. [Narrating Life Images of a Self-made Senior Citizen Through Life Review]. *J. Gerontechnolo. Serv. Manag.* 5, 265–283. doi: 10.6283/jocsg.201711_5(3)0.265

Cuijpers, P., Vogelzangs, N., Twisk, J., Kleiboer, A., Li, J., and Penninx, B. W. (2013). Differential mortality rates in major and subthreshold depression: meta-analysis of studies that measured both. *Br. J. Psychiatry* 202, 22–27. doi: 10.1192/bjp.bp.112.112169

Demoulin, K. C. N. (2016). *Marketing Research with SAS Enterprise Guide*. New York, NY: Routledge.

Derdaele, E., Toussaint, L., Thauvoye, E., and Dezutter, J. (2019). Forgiveness and late life functioning: the mediating role of finding ego-integrity. *Aging Ment. Health* 23, 238–245. doi: 10.1080/13607863.2017.1399346

Dezutter, J., Toussaint, L., and Leijssen, M. (2016). Forgiveness, ego-integrity, and depressive symptoms in community-dwelling and residential elderly adults. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 71, 786–797. doi: 10.1093/geronb/gbu146

Domino, G., and Affonso, D. D. (1990). A personality measure of Erikson's life stages: the inventory of psychosocial balance. *J. Pers. Assess.* 54, 576–588. doi: 10.1080/00223891.1990.9674021

Epstein, J., Santo, R. M., and Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J. Clin. Epidemiol.* 68, 435–441. doi: 10.1016/j.jclinepi.2014.11.021

Erikson, E. H. (1963). *Childhood and Society*. New York, NY: Norton.

Erikson, E. H. (1980). *Identity and the Life Cycle*. New York, NY: Norton.

Erikson, E. H. (1982). *The Life Cycle Completed*. New York, NY:: Norton.

Finney, G. R., Minagar, A., and Heilman, K. M. (2016). Assessment of mental status. *Neurol. Clin.* 34, 1–16. doi: 10.1016/j.ncl.2015.08.001

Galea, M., and Woodward, M. (2005). Mini-mental state examination (MMSE). *Aust. J. Physiother.* 51:198. doi: 10.1016/s0004-9514(05)70034-9

Gregory, R. J. (2004). *Psychological Testing: History, Principles, and Applications*. Needham Heights, MA: Allyn & Bacon.

Hannah, M. T., Domino, G., Figueredo, A. J., and Hendrickson, R. (1996). The prediction of ego integrity in older persons. *Educ. Psychol. Meas.* 56, 930–950. doi: 10.1177/0013164496056006002

Kaiser, H. F., and Rice, J. (1974). Little Jiffy, Mark Iv. *Educ. Psychol. Meas.* 34, 111–117. doi: 10.1177/001316447403400115

Kleijn, G., Post, L., Witte, B. I., Bohlmeijer, E. T., Westerhof, G. J., Cuijpers, P., et al. (2016). Psychometric characteristics of a patient reported outcome measure on ego-integrity and despair among cancer patients. *PLoS One* 11:e0156003. doi: 10.1371/journal.pone.0156003

Lavrakas, P. (2008). *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage Publications, doi: 10.4135/9781412963947

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Bloomington, IN: Indiana University.

Lu, J.-F. R., Tseng, H.-M., and Tsai, Y.-J. (2003). Assessment of health-related quality of life in Taiwan (I): development and psychometric testing of SF-36 Taiwan Version. *Taiwan J. Public Health* 22, 501–511.

Min, S. (2020). A study on the ego-integrity. *Int. J. Psychos. Rehabil.* 24, 1643–1650. doi: 10.37200/IJPR/V24I7/PR270147

Munro, B. H. (2005). *Statistical Methods for Health Care Research*, 5th Edn. New York: Lippincott Williams & Wilkins.

Olmedo Moreno, E. M., de Luna, E. B., Olmos Gómez, M. D. C., and López, J. E. (2014). Structural Equations Model (SEM) of a questionnaire on the evaluation of intercultural secondary education classrooms. *Suma Psicol.* 21, 107–115. doi: 10.1016/S0121-4381(14)70013-X

Polit, D., and Beck, C. T. (2004). *Nursing Research: Principles and Methods*, 7th Edn. New York, NY: Lippincott Williams &Wilkins.

Rived-Ocaña, M., Schweer-Collins, M. L., Rodríguez-González, M., Crabtree, S. A., del Cid, L. B.-G., and Hargrave, T. D. (2020). Spanish adaptation of the

relational ethics scale. *Contemp. Fam. Ther.* 42, 346–359. doi: 10.1007/s10591-020-09540-y
Saint Onge, J. M., Krueger, P. M., and Rogers, R. G. (2014). The relationship between major depression and nonsuicide mortality for U.S. adults: the importance of health behaviors. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 69, 622–632. doi: 10.1093/geronb/gbu009
Sharma, S., Mukherjee, S., Kumar, A., and Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *J. Bus. Res.* 58, 935–943. doi: 10.1016/j.jbusres.2003.10.007
Sherbourne, and Ware, J. (1992). The MOS 36-item short-form health survey (SF-36). *Med. Care* 30, 473–483. doi: 10.1097/00005650-199206000-00002
Taber, K. S. (2018). The use of cronbach's alpha when developing and reporting research instruments in science education. *Res. Sci. Educ.* 48, 1273–1296. doi: 10.1007/s11165-016-9602-2
Tseng, H. M., Lu, J. F., and Gandek, B. (2003). Cultural issues in using the SF-36 Health Survey in Asia: results from Taiwan. *Health Qual. Life Outcomes* 1:72. doi: 10.1186/1477-7525-1-72
Üzar-Özçetin, Y. S., and Ercan-Şahin, N. (2020). Descriptive phenomenological study on ego-integrity among older people in nursing homes. *Nurs. Health Sci.* 22, 472–479. doi: 10.1111/nhs.12715
Van Hiel, A., and Vansteenkiste, M. (2009). Ambitions fulfilled? The effects of intrinsic and extrinsic goal attainment on older adults' ego-integrity and death attitudes. *Int. J. Aging Hum. Dev.* 68, 27–51. doi: 10.2190/AG.68.1.b
Westerhof, G. J., Bohlmeijer, E. T., and McAdams, D. P. (2017). The relation of ego integrity and despair to personality traits and mental health. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 72, 400–407. doi: 10.1093/geronb/gbv062
Williams, B., Onsman, A., and Brown, T. (2010). Exploratory factor analysis: a five-step guide for novices. *Austr. J. Paramed.* 8, 1–13. doi: 10.33151/ajp.8.3.93

Conflict of Interest: The authors declare the study received funding from China Medical University Hospital (Grant No. DMR-106-147). The funding body played a role in the design of the study and in writing the manuscript.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 February 2022
doi: 10.3389/fpsyg.2021.771095

Check for updates

Validation and Psychometric Properties of the Minnesota Living With Heart Failure Questionnaire in Individuals With Coronary Artery Disease in Lithuania

Julija Gecaite-Stonciene[1]*, Julius Burkauskas[1], Adomas Bunevicius[1], Vesta Steibliene[1], Jurate Macijauskiene[2], Julija Brozaitiene[1], Narseta Mickuviene[1] and Nijole Kazukauskiene[1]

[1] Laboratory of Behavioral Medicine, Neuroscience Institute, Lithuanian University of Health Sciences, Palanga, Lithuania, [2] Faculty of Nursing, Lithuanian University of Health Sciences, Kaunas, Lithuania

OPEN ACCESS

Edited by:
Victor Zaia,
Faculdade de Medicina do ABC, Brazil

Reviewed by:
Erik Montagna,
Centro Universitário FMABC, Brazil
Erica Hokama,
Universidade São Judas Tadeu, Brazil
Pedro Afonso Cortez,
Universidade Metodista de São Paulo (UMESP), Brazil

****Correspondence:***
Julija Gecaite-Stonciene
julija.gecaite@lsmuni.lt

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *05 September 2021*
Accepted: *21 December 2021*
Published: *04 February 2022*

Citation:
Gecaite-Stonciene J, Burkauskas J, Bunevicius A, Steibliene V, Macijauskiene J, Brozaitiene J, Mickuviene N and Kazukauskiene N (2022) Validation and Psychometric Properties of the Minnesota Living With Heart Failure Questionnaire in Individuals With Coronary Artery Disease in Lithuania. Front. Psychol. 12:771095. doi: 10.3389/fpsyg.2021.771095

Background: Health-related quality of life (HRQoL) is known to be impaired in individuals with coronary artery disease (CAD), especially in those after a recent acute coronary syndrome (ACS). Heart failure (HF) is a common burden in this population that significantly contributes to worsening HRQoL. To accurately measure the level of HRQoL in individuals with CAD after ACS, disease-specific scales, such as the Minnesota living with heart failure questionnaire (MLHFQ), are recommended. Nevertheless, to date, there has not been a study that would comprehensively evaluate the psychometric properties of the MLHFQ in a large sample of individuals with CAD after ACS. The debate regarding the internal structure of MLHFQ is also still present. Hence, this study aimed to translate the MLHFQ and evaluate its internal structure, reliability/precision, and validity in individuals with CAD following ACS in Lithuania.

Methods: In the cross-sectional study, 1,083 participants (70% men, age M = 58, SD = 9) were evaluated for sociodemographic and clinical characteristics. HRQoL was measured using the MLHFQ and the Short Form-36 health survey (SF-36). In addition, exercise capacity (EC) was also evaluated in the study patients, using a standardized computer-driven bicycle ergometer.

Results: The internal consistency of the MLHFQ subscales (0.79–0.88) was found to be good. Confirmatory factor analysis (CFA) provided the support for the three-factor model ("physical domain," "social domain," and "emotional domain") of the MLHFQ and showed acceptable fit [comparative fit indices (CFI) = 0.894; goodness-of-fit (GFI) = 0.898; non-normal fit index (NFI) = 0.879, and root mean square error of approximation (RMSEA) = 0.073]. Regarding convergent evidence, significant associations were found between the MLHFQ domains and the SF-36 domains and EC (r's range 0.11–0.58).

Conclusion: The current study completed cultural validation and provided further information on the psychometric characteristics of the MLHFQ in Lithuania, suggesting MLHFQ as a valid and reliable instrument to measure HRQoL. The Lithuanian version of MLHFQ is best described by a three-factor solution, measuring physical, social, and emotional dimensions of HRQoL among individuals with CAD following ACS.

Keywords: Minnesota living with heart failure questionnaire, quality of life, factorial structure, measures, psychometrics, validation, coronary artery disease, cross-cultural

INTRODUCTION

Coronary artery disease (CAD) is considered the leading cause of morbidity and mortality worldwide and is a major factor in the development and progression of heart failure (HF) (Mozaffarian et al., 2016). In the developed countries, approximately 1–2% of adults meet the criteria for HF; however, in people who are 70 years of age or older, the prevalence is ≥10% (Mosterd and Hoes, 2007). Given the aging population in Europe, the prevalence of HF is likely to rise (Iacoviello and Antoncecchi, 2013). Acute coronary syndrome (ACS), such as myocardial infarction (MI) and unstable angina pectoris (Amsterdam et al., 2014), is one of the most common presentations of CAD, leading to the increase of long-term incidence of HF up to 30% (Qureshi et al., 2018; Cordero et al., 2021). Moreover, one out of ten individuals after ACS that is considered in the category of a low-risk for development of HF, eventually develops a new-onset of HF (Cordero et al., 2021). Thus, even though they are distinct conditions, CAD, ACS, and HF often coexist together.

Health-related quality of life (HRQoL) in individuals with heart-related conditions is known to be impaired (Callus et al., 2020; Journiac et al., 2020). HRQoL is especially problematic in those with CAD after ACS with HF (Staniute et al., 2015; Kazukauskiene et al., 2019). According to the WHO, QoL is described as the perception of individuals' position in life in the context of the culture and value systems in which they live and in relation to their objectives, expectations, concerns, and standards. In clinical settings, HRQoL has been described as the person's perception of the influence that the illness has on their life. Specifically, HRQoL measures the person's perception about living with the disease through functional capacity, occupational health, the general perception of health status, and psychological/social functioning in the context in which they operate (Haas, 1999; Erceg et al., 2013; Falk et al., 2013; Mogle et al., 2017).

To evaluate the HRQoL of an individual, generic and disease-specific instruments are commonly used. One of the most known disease-specific instruments for measuring HRQoL is the Minnesota Living with HF Questionnaire (MLHFQ) (Rector and Cohn, 1992; Garin et al., 2014). The MLHFQ has advantages over generic scales, as it has the responsiveness and the capacity to discriminate between different magnitudes of change in individuals' HRQoL (Garin et al., 2009; Rajati et al., 2016; Napier et al., 2018; Gonzalez-SaenzdeTejada et al., 2019). This scale has been adapted and translated to at least 34 languages in various countries and has shown good psychometric properties (Sneed et al., 2001; Bennett et al., 2003; Heo et al., 2005; Ho et al., 2007; Garin et al., 2008, 2013, 2014; Moon et al., 2012; Bilbao et al., 2016; Zahwe et al., 2020). Yet, up to date, there is a lack of studies that would examine the reliability/precision and validity of the MLHFQ in a large sample of individuals with CAD following ACS. To our knowledge, there are no psychometric studies on the MLHFQ in the Lithuanian population as well.

Furthermore, with regard to the psychometric properties of the MLHFQ, studies that investigated its internal structure yielded inconsistent results. Specifically, some of the studies indicated a two-factor solution (i.e., physical and emotional dimensions) (Middel et al., 2001; Garin et al., 2013; Brokalaki et al., 2015; Bilbao et al., 2016) that includes the original MLHFQ study (Rector and Cohn, 1992). However, ample published studies that validated the three-factor internal structure of the MLHFQ also exist (Heo et al., 2005; Ho et al., 2007; Moon et al., 2012; Lambrinou et al., 2013; Munyombwe et al., 2014; Mogle et al., 2017; Barnett et al., 2019; Zahwe et al., 2020) that report the potential existence of a third factor of "social domain." Even though the inconsistent findings might be attributed to a variety of methodological aspects and cultural differences, further studies in a large sample of individuals with HF are warranted.

Considering the knowledge gap in terms of psychometric properties of the MLHFQ, this study aimed to translate and evaluate the applicability, internal consistency, and validity of the MLHFQ and to investigate the internal structure when administered on individuals with CAD following ACS in Lithuania.

MATERIALS AND METHODS

Study Procedure

In sum, 1,190 consecutive patients with CAD were invited to participate in the larger study during the period from February 2014 to January 2019. The inclusion criteria were (1) current diagnosis of ACS, as defined by acute MI or unstable angina pectoris, and (2) the participation in the cardiac rehabilitation program at Lithuanian University of Health Sciences, Neuroscience Institute, Hospital Palangos Klinika within 1–2 weeks following treatment for ACS.

The exclusion criteria were (1) unstable cardiovascular condition (n = 52), (2) other severe illness (e.g., kidney failure or musculoskeletal pathology) (n = 28), and (3) unwillingness to participate in the study (n = 27). In sum, the final sample was comprised of 1,083 subjects (76% men, age M = 57, SD = 9). All participants underwent standardized diagnostic and treatment procedures for secondary CAD prevention, based on the established guidelines (Gibbons et al., 2002; Piepoli et al., 2010; Fletcher et al., 2013; O'Gara et al., 2013).

Within 2 days of admission to the cardiac rehabilitation, all study participants were assessed for demographic (i.e., age, gender, education, and marital status) and clinical factors [i.e., New York Heart Association (NYHA) functional class, angina pectoris class, medical diagnosis, and body mass index (BMI)]. Standard echocardiography testing was performed to evaluate left ventricular ejection fraction (LVEF).

All study participants independently completed the self-report questionnaires. In terms of HRQoL scales, we used the MLHFQ (Rector and Cohn, 1992) and the 36-Item Short Form Medical Outcome Questionnaire (SF-36) (Ware and Sherbourne, 1992).

All procedures conducted in current research involving human subjects were in compliance with the ethical principles of the Biomedical Research Ethics Committee for Biomedical Research at Lithuanian University of Health Sciences and conformed to the principles outlined in the Declaration of Helsinki. Informed consent was attained from all individual participants included in the study.

Measures

The Minnesota Living With Heart Failure Questionnaire

The MLHFQ is comprised of 21 items evaluating physical, social, and emotional aspects of life, referring to weaknesses often associated with the cardiac insufficiency profile (Rector and Cohn, 1992). The major advantage of the MLHFQ for those with HF, in comparison to other generic HRQoL assessment scales, is that it specifically addresses the daily living challenges and context of those experiencing HF.

The questions are based on a six-point Likert scale from 0 ("no impact of HF on HRQoL") to 5 (the significant negative impact of HF on HRQoL). The MLHFQ has a total score (21 items, score range 0–105), the sum of points of the physical dimension subscale (8 items, score range 0–40), and the sum of points of the emotional dimension subscale (5 items, scores range 0–25). Higher scores indicate worse HRQoL.

Permission to translate the original MLHFQ into a Lithuanian version was obtained from the University of Minnesota, which holds its copyright. The questionnaire was translated by the two independent bilingual translators. The third bilingual translator without any previous knowledge about the MLHFQ back-translated and re-conciliated the Lithuanian version. The final version was completed after the revisions were provided by several different experts in the field that were fluent in both languages, i.e., the cardiologist (Julija Brozaitiene), the medical psychologists (Julija Gecaite-Stonciense and Julius Burkauskas), and the registered nurse (Nijole Kazukauskiene).

36-Item Short Form Medical Outcome Questionnaire

The SF-36 is comprised of 8 multi-item scales that evaluate HRQoL on 8 domains: (1) physical functioning, (2) social functioning, role limitations due to (3) emotional problems and (4) physical problems, (5) mental health, (6) energy/vitality, (7) pain, and (8) general health perception. Each SF-36 domain is scored from 0 to 100, where higher scores reflect better HRQoL (Ware and Sherbourne, 1992). Internal-consistency reliability (Cronbach's α) of eight subscales has been found to range between 0.57 and 0.85.

Exercise Capacity Testing

In addition, study participants also completed testing for EC. Participants' (EC) was measured by the cardiologist (Julija Brozaitiene) using a standardized computer-driven bicycle ergometer with workload rising by 25 watts (W) every 3 min (Fletcher et al., 2013). The peak of workload (PW) in watt (W) or metabolic equivalent of task (MET) (1 MET = 3.5 ml of oxygen uptake per kilogram of body weight per min) at the completion of the test reflected EC. Detailed procedures of EC have been reported in our study elsewhere (Kazukauskiene and Burkauskas, 2018). Reduced EC was regarded as $\leq$50 W, as this cut score is associated with moderate-to-severe functional impairment and cardiac symptoms (Gibbons et al., 2002; Piepoli et al., 2010; Fletcher et al., 2013).

Statistical Analysis

Statistical analysis was performed using the Statistical Package for Social Sciences, SPSS Statistics for Windows, Version 22.0.0.0 (IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY, United States: IBM Corp). Data are expressed as a mean ± SD for continuous variables and as a number (percent) for qualitative variables. The distribution of measures was assessed using skewness and kurtosis analysis. In a total sample, the scores were approximately normally distributed for the total score and the physical subscale with skewness of 0.518 (SE = 0.325) and kurtosis of − 0.280 (SE = 0.149) for the total score and skewness of 0.565 (SE = 0.341) and kurtosis − 0.496 (SE = 0.264) for the physical subscale. Values of the emotional subscale were 1.133 (SE = 0.074) for skewness and 0.893 (SE = 0.27) for kurtosis, respectively. Taking into account the large sample size of this current study, the latter scale was considered to be normally distributed (Kim, 2013).

Cronbach's α coefficients were calculated to evaluate the internal-consistency reliability of the MLHFQ and its dimensions and to contrast these with the SF-36 scales and composites. For each of the subscales of the MLHFQ and SF-36 domains, the floor and ceiling effects were described as the proportion of individuals at each subscale, who achieved the lowest or the highest possible score. The floor and ceiling effects were considered to be present when at least 15% of respondents received the lowest or the highest rating, respectively.

Convergent evidence was evaluated by measuring the Pearson Correlation Coefficients of the MLHFQ and its dimensions and all the scores of the SF-36 subscales. It was expected that the MLHFQ total score would be highly associated with the SF-36 vitality and social functioning, while scores of the MLHFQ dimension pertaining to physical health would highly correlate with the SF-36 physical functioning, vitality, and social functioning. Further, it was expected that the MLHFQ dimension pertaining to emotional health would be highly correlated with the SF-36 emotional wellbeing and vitality, respectively.

Construct validity of the MLHFQ was evaluated by comparing the MLHFQ and its scores on a different dimension to age (<65 years vs. ≥ 65 years), gender (male, female), NYHA functional class (NYHA I–II class vs. NYHA III–IV class), LVEF (LVEF $\leq$ 40% vs. LVEF $>$40%), EC ($\leq$50 W vs. $>$50 W), and BMI (<30 kg/m^2 vs. ≥ 30 kg/m^2) using the Independent Samples *t*-test.

Exploratory factor analyses (EFA) (sub—sample 1, n = 541), sample adequacy, and factorability of data were analyzed using the Kaiser-Meyer-Olkin (KMO) measure and the Bartlett test for sphericity. We performed an initial EFA to determine the dimensionality of the MLHFQ questionnaire using principal component analysis with the varimax rotation method. Cut scores for factor loadings were set at 0.4. Items cross-loading on multiple factors were assigned to the factor with higher loading. We used confirmatory factor analysis (CFA) (sub—sample 2, n = 542) to determine the internal structure of the MLHFQ using maximum likelihood (ML) estimation. The fitness of the model with the data was measured by calculating the absolute and comparative fit indices (CFI). Absolute fit indices included chi-square goodness-of-fit (GFI), non-normal fit index (NFI), and root mean square error of approximation (RMSEA).

RESULTS

In terms of descriptive information, **Table 1** represents the basic sociodemographic characteristics of all individuals participating in the study. In summary, the majority of study participants (79%) had moderate HF symptoms (NYHA II functional class), 14% had severe HF symptoms, and were assigned to NYHA III–IV functional class. The minority of the participants (7%) had mild HF symptoms (NYHA I functional class). In sum, 28% of study participants had angina pectoris, 61% had acute MI, and 11% had previous MI. Eighty-nine percent of individuals were treated with beta-blockers, 94% – statins, 81% – angiotensin-converting enzyme inhibitors, and 15% – diuretics, and other medications.

Mean scores of the MLHFQ and the SF-36 domains are shown in **Table 2**. The lowest SF-36 score was for the role limitations due to physical problems subscales, while the highest score was for the subscale of emotional wellbeing. Cronbach's coefficients α were greater than 0.70 for all subscales, except the subscales of general health (0.688) and social functioning (0.572). In the current study, the MLHFQ had adequate internal-consistency reliability (Cronbach's α of Total score = 0.91, Cronbach's α of physical dimension = 0.88, and emotional dimension = 0.82).

Table 2 also shows the ceiling ("poorest" HRQoL) and floor ("best" HRQoL) effects of the MLHFQ. The ceiling effect was present for the total score of the MLHFQ 2.4% (n = 26), physical dimension 5.7% (n = 62), and emotional dimension 14.6% (n = 158). The floor effect was present for the physical dimension 0.2% (n = 2) and emotional dimension of the MLHFQ 3% (n = 3). In contrast, the SF-36 ceiling represents "best" HRQoL, while floor indicates "poorest" HRQoL effects. The ceiling effect was present for the SF-36 subscales of physical functioning 3% (n = 3), role limitation due to physical problems 51.4% (n = 557), role limitations due to emotional problems 33.5% (n = 363), social functioning 6%, (n = 7), vitality 2% (n = 2), pain 5.7%, (n = 62), and general health 5% (n = 5). Floor effect was present for the SF-36 subscales of physical functioning 3.4% (n = 37), role limitation due to physical problems 14.1% (n = 153), role limitations due to emotional problems 40.4% (n = 437), social functioning 21.3%, (n = 231), emotional wellbeing 4.2% (n = 45), vitality 1.8% (n = 20), pain 9.5%, (n = 103), and general health 6% (n = 6).

As demonstrated in **Table 3**, the MLHFQ total score was mostly associated with the SF-36 vitality (r = − 0.597, $p < 0.001$) and the SF-36 social functioning (r = − 0.594, $p < 0.001$), while scores on the MLHFQ dimension pertaining to physical health were mostly associated with the SF-36 physical functioning (r = − 0.571, $p < 0.001$), SF-36 social functioning (r = − 0.577, $p < 0.001$), and SF-36 vitality (r = − 0.593, $p < 0.001$). The MLHFQ dimension pertaining to emotional health was mostly associated with the SF-36 emotional wellbeing

TABLE 1 | Characteristics of all study patients.

	Total group
	***n* = 1083**
Age, years (mean ± SD)	57.44 ± 9.03
Gender, *n* (%)	
Male	821 (75.8)
Female	262 (24.2)
Marital status, *n* (%)	
Cohabiting	904 (83.5)
Single	21 (1.9)
Divorced	79 (7.3)
Widowed	79 (7.3)
Education, *n* (%)	
Up to 8 years	85 (7.8)
High school graduate	536 (49.5)
College/university degree	462 (42.7)
Body mass index ≥ 30 kg/m^2, *n* (%)	484 (44.7)
Medical diagnosis, *n* (%)	
Angina pectoris	302 (27.9)
Acute myocardial infarction	661 (61.0)
Previous myocardial infarction	120 (11.1)
New York Heart Association functional class, *n* (%)	
I	74 (6.8)
II	859 (79.3)
III-IV	150 (13.9)
Left ventricular ejection fraction, *n* (%)	
≤40%	121 (11.2)
>40%	962 (88.8)
Exercise capacity, *n* (%)	
≤50 W	437 (40.4)
>50 W	646 (59.6)
Medications, *n* (%):	
Nitrates	270 (24.9)
Angiotensin-converting-enzyme inhibitors	876 (80.9)
Beta-blockers	964 (89.0)
Diuretics	158 (14.6)
Statines	1015 (93.7)
Benzodiazepines	173 (16.0)

TABLE 2 | Health-related quality of life scales characteristics.

HRQoL measure	No. of items	Mean ± SD	Median [IQR]	Min	Max	Ceiling, *n* (%)	Floor, *n* (%)	Cronbach α
MLHFQ			*Score*		*Score*			
Total score	21	31.33 ± 19.84	29 (16–45)	0	100	26 (2.4)	0	0.912
Physical dimension	8	12.89 ± 9.28	12 (5–20)	0	40	62 (5.7)	2 (0.2)	0.885
Emotional dimension	5	5.72 ± 5.29	4 (2–9)	0	25	158 (14.6)	3 (0.3)	0.828
SF-36								
Physical functioning	10	67.75 ± 20.16	70 (55–85)	0	100	3 (0.3)	37 (3.4)	0.859
Role limitations due To physical problems	4	29.18 ± 36.87	0 (0–50)	0	100	557 (51.4)	153 (14.1)	0.829
Role limitations due to emotional problems	3	52.85 ± 43.74	66.67 (0–100)	0	100	363 (33.5)	437 (40.4)	0.851
Social functioning	2	66.66 ± 26.47	66.67 (44.44–88.89)	0	100	7 (0.6)	231 (21.3)	0.572
Emotional well-being	5	68.13 ± 19.33	72 (56–84)	4	100	0	45 (4.2)	0.800
Vitality	4	58.36 ± 21.02	60 (45–75)	0	100	2 (0.2)	20 (1.8)	0.718
Pain	2	50.35 ± 27.62	44.44 (33.33–66.67)	0	100	62 (5.7)	103 (9.5)	0.752
General health	5	52.89 ± 18.69	50 (40–65)	0	100	5 (0.5)	6 (0.6)	0.688

HRQoLf, health-related quality of life; IQR, interquartile range; min, lowest; max, highest; MLHFQ, Minnesota Living with Heart Failure Questionnaire; SF-36, Medical Outcomes Study 36-Item Short Form Health Survey; α, Cronbach's alpha coefficient.

($r = -0.646$, $p < 0.001$) and the SF-36 vitality ($r = -0.576$, $p < 0.001$), respectively.

The strong associations were observed between theoretically similar domains, for example, the MLHFQ physical dimension and the SF-36 measures of physical HRQoL, and between the MLHFQ emotional dimension and the SF-36 measures of mental HRQoL, providing the evidence for convergent evidence (**Table 3**).

Table 4 shows the EFA of the MLHFQ items. To determine the internal structure, all 21 items of the MLHFQ underwent an EFA with data of sub−sample 1 ($n = 541$). A value of 0.926 on the KMO test indicated adequate correlation matrices. The Bartlett sphericity test was significant at χ^2 = 4882,855 ($p < 0.001$), indicating the presence of significant correlations and reinforcing the relevance of the factor analysis. In all three factors, the multiple loadings of items had factor-loading values of >0.40. Factor analysis with three-factor explained 52.6% of the total variance, out of which 36.7% was explained by the first factor, 8.8% by the second factor, and 7.1% by the third factor. Items 1–6, 12, and 13 loaded heavily onto the first factor. Items 16–21 loaded on to the second factor. Items 7–11 and 14 loaded heavily onto the third factor. Item 15 did not load on any of the subscales or the overall factor and was eliminated from the CFA analysis. CFA using data of sub−sample 2 ($n = 542$) confirmed the EFA solution. CFA showed acceptable fit [CFI = 0.895; GFI = 0.882; NFI = 0.864, RMSEA = 0.074 95% CI = (0.068;0.080)].

TABLE 3 | Convergent evidence of the Minnesota Living with Heart Failure Questionnaire (MLHFQ), 36-Item Short Form Health Survey (SF-36), and exercise capacity in the overall sample.

	MLHFQ		
	Total score	Physical dimension	Emotional dimension
SF-36			
Physical functioning	−0.520 (<0.001)	−0.571 (<0.001)	−0.402 (<0.001)
Role limitations due to physical problems	−0.402 (<0.001)	−0.387 (<0.001)	−0.267 (<0.001)
Role limitations due to emotional problems	−0.375 (<0.001)	−0.323 (<0.001)	−0.372 (<0.001)
Social functioning	−0.594 (<0.001)	−0.577 (<0.001)	−0.451 (<0.001)
Emotional well-being	−0.524 (<0.001)	−0.445 (<0.001)	−0.646 (<0.001)
Vitality	−0.597 (<0.001)	−0.593 (<0.001)	−0.576 (<0.001)
Pain	−0.454 (<0.001)	−0.460 (<0.001)	−0.319 (<0.001)
General health	−0.516 (<0.001)	−0.469 (<0.001)	−0.476 (<0.001)
Exercise capacity	−0.248 (0.000)	−0.289 (0.000)	−0.208 (0.000)

SF-36, Medical Outcomes Study 36-Item Short Form Health Survey.

Table 5 provides Pearson Correlation Coefficients between the factors of the MLHFQ, the SF-36, and EC. As expected, the scores of the factors were significantly correlated with the SF-36 and EC, respectively, supporting convergent evidence.

In younger women, the higher NYHA functional class and reduced EC were associated with the higher MLHFQ scores, indicating worse HRQoL. Individuals younger than 65 years scored significantly higher only on the MLHFQ total score ($p < 0.05$, with a moderate effect size). Female gender, higher NYHA functional class, obesity (BMI ≥ 30 kg/m^2), and reduced EC were linked with higher scores on all dimensions of the MLHFQ ($p < 0.001$, with a small-to-moderate effect size), indicating individuals' worse HRQoL. The comparison analyses showed no differences between LVEF $\leq$ 40% vs. LVEF $\leq$40% groups (**Figure 1**).

DISCUSSION

Our results indicated that the MLHFQ could be considered as a valid and reliable instrument for the evaluation of HRQoL in individuals with CAD following ACS during a cardiac

TABLE 4 | Factor analysis of the MLHFQ items.

	Item	Factor 1 (physical)	Factor 2 (emotional)	Factor 3 (social)
1	Swelling in your ankles, legs	0.434	0.183	−0.084
2	Resting during day	0.720	0.140	0.286
3	Walking or climbing stairs difficult	0.765	0.115	0.297
4	Working around house difficult	0.602	0.111	0.548
5	Away from home difficult	0.647	0.100	0.444
6	Sleeping difficult	0.539	0.312	0.157
7	Relating to or doing things with friends	0.419	0.218	0.545
8	Working to earn a living difficult	0.199	0.062	0.670
9	Recreational activities difficult	0.373	0.088	0.726
10	Sexual activities difficult	0.003	0.064	0.734
11	Eating less foods, I like	0.126	0.258	0.508
12	Shortness of breath	0.651	0.212	−0.009
13	Fatigue	0.724	0.366	0.180
14	Hospitalization	0.001	0.304	0.516
15*	Medical costs	0.396	0.378	0.246
16	Side effects from medications	0.193	0.529	0.173
17	Feeling burden to family or friends	0.011	0.479	0.401
18	Feeling loss of self-control	0.160	0.773	0.126
19	Being worried	0.239	0.753	0.142
20	Difficulty concentrating or remembering	0.418	0.682	0.043
21	Being depressed	0.268	0.767	0.157
Cronbach's α		0.864	0.863	0.827

Data refer to subsample 1 of sample 2.
**Item not belonging to any factor.*
MLHFQ, Minnesota Living with Heart Failure Questionnaire.

TABLE 5 | Convergent evidence: Pearson correlation coefficients between the factors of Minnesota Living with Heart Failure Questionnaire (MLHFQ), 36-Item Short Form Health Survey (SF-36), and exercise capacity in the overall sample.

	Factors		
	1	2	3
SF-36			
Physical functioning	−0.575 (<0.001)	−0.331 (<0.001)	−0.406 (<0.001)
Role limitation due to physical problems	−0.379 (<0.001)	−0.355 (<0.001)	−0.271 (<0.001)
Role limitation due to emotional problem	−0.324 (<0.001)	−0.284 (<0.001)	−0.367 (<0.001)
Social functioning	−0.544 (<0.001)	−0.514 (<0.001)	−0.460 (<0.001)
Mental health	−0.442 (<0.001)	−0.323 (<0.001)	−0.622 (<0.001)
Energy/vitality	−0.594 (<0.001)	−0.370 (<0.001)	−0.571 (<0.001)
Bodily pain	−0.458 (<0.001)	−0.357 (<0.001)	−0.320 (<0.001)
General health	−0.475 (<0.001)	−0.355 (<0.001)	−0.486 (<0.001)
Exercise capacity	−0.294 (<0.001)	−0.112 (<0.001)	−0.215 (<0.001)

MLHFQ, Minnesota Living with Heart Failure Questionnaire; SF-36, medical outcomes study 36-item short form health survey;
Factor 1, physical dimension.
Factor 2, social dimension;
Factor 3, emotional dimension.

rehabilitation program in Lithuania. The current data support sound psychometric properties of MLHFQ and its response to therapeutic interventions. The current findings also suggest the existence of the third factor in the MLHFQ in individuals with CAD following ACS, some of whom lack HF symptoms.

To the best of our knowledge, this is the first study assessing the performance of the MLHFQ among individuals with CAD undergoing ACS and is among the very few (Garin et al., 2008, 2013) which have included a comparison with the SF-36.

The MLHFQ's three-factor internal-consistency reliability was good, which is in line with the performance of this instrument in randomized clinical trials performed in individuals with HF (Heo et al., 2005; Ho et al., 2007). For the distinct physical and emotional factors described in the original MLHFQ validation, empirical support was found (Rector and Cohn, 1992).

In addition, it is important to note that in our study the MLHFQ had relatively mild ceiling and floor effects, which are important psychometric properties of any scale (Feeny et al., 2013). Based on the current results, the MLHFQ can be considered as an applicable instrument to measure relatively minor changes in the self-reported health status of individuals with CAD after ACS representing symptoms of HF. The required threshold rate for floor and ceiling effects was not reached for the MLHFQ and their dimensions suggesting a good sensitivity of the MLHFQ for evaluating perceived HRQoL of individuals with CAD who underwent ACS.

Further, the strong correlations between theoretically related dimensions of the MLHFQ and the SF-36 have provided evidence of convergent evidence; results that also give the support for the underlying constructs supporting the instrument's criterion validity in this population (Boateng et al., 2018). The MLHFQ physical and emotional dimensions were strongly associated with similar domains of the SF-36 measures of physical and mental HRQoL (Garin et al., 2008, 2013). In terms of objective measures, the MLHFQ was associated with EC, especially the physical dimension of MLHFQ and EC, indicating that EC plays a major role in decreasing HRQoL in those individuals (Bussoni et al., 2013).

Previous studies have suggested that the functional classification of NYHA was characterized by the severity of the symptoms of HF. Thus, it was expected that the MLHFQ would be useful for individuals with different HF severity levels (Carvalho et al., 2009). In the current study, the NYHA classification and the value of EC measured as external work in watts were consistent with higher MLHFQ scores, indicating worse HRQoL. A significant correlation between the NYHA functional classification and the score for the Spanish version

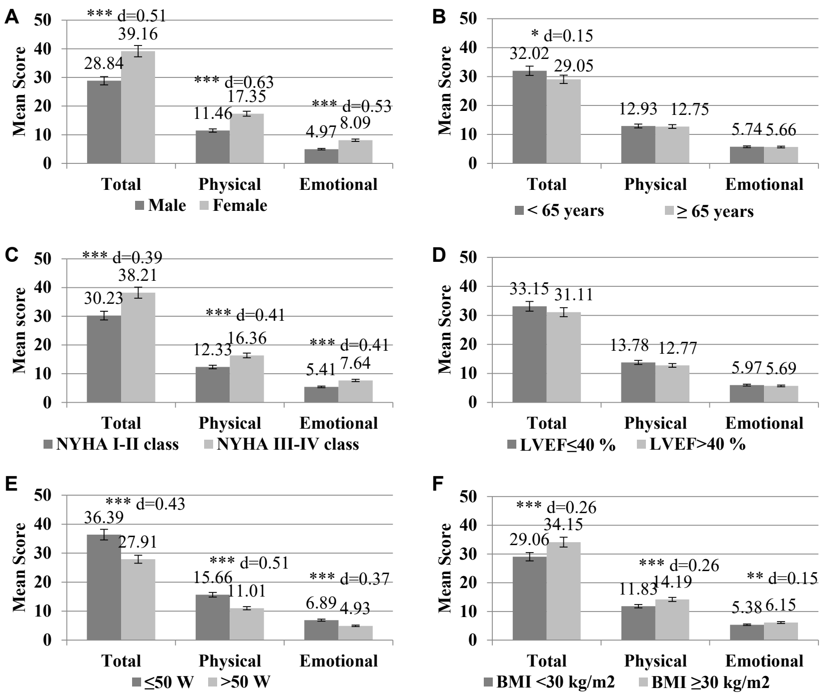

FIGURE 1 | The differences of mean scores of the Minnesota living with heart failure questionnaire (MLHFQ) dimensions (global, physical, and emotional health) according to gender **(A)**, age **(B)**, New York Heart Association, NYHA functional class **(C)**, left ventricular ejection fraction, LVEF **(D)**, watts for exercise capacity **(E)**, and body mass index and BMI **(F)**. Effect size **(D)** for the difference between the groups. (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; d, Cohen's d effect size).

of the MLHFQ was recorded in an earlier review (Parajon et al., 2004). Parajon et al. (2004) found that higher scores on the MLHFQ, suggesting lower HRQoL, were correlated with higher NYHA functional class and hospital admissions over the previous year. In a small sample of individuals with advanced chronic HF, Morcillo et al. (2007) reported strong associations between the MLHFQ, a functional class, and the SF-36. The findings were similar in our study as well.

Further, our study suggests the presence of a third "social" dimension of the MLHFQ. Among the other three-factor models, the best results obtained with this model were proposed by Garin et al. (2013). With regard to the internal structure, our study is also in line with other previous reports where HF was mostly a primary condition (Heo et al., 2005; Ho et al., 2007; Moon et al., 2012; Lambrinou et al., 2013; Munyombwe et al., 2014; Mogle et al., 2017; Barnett et al., 2019; Zahwe et al., 2020). Some of those studies (Heo et al., 2007; Mogle et al., 2017; Zahwe et al., 2020) had issues that fully support the psychometric soundness of several items. In our case, similar to Mogle et al. (2017) and Zahwe et al. (2020) reports, item 15 ("medical costs") was problematic and did not fall under any of the factors. We assume that this issue can be partially attributed to cultural socioeconomic differences across countries, such as variations in healthcare policies across nations, which, depending on whether the country has a universal healthcare system, makes this item more or less important.

In addition, in the current study, sample disequilibrium was observed, as the majority (75.8%) of the participants were men with a mean age of 57.4. It is well established that CAD is more prevalent among men than women (Khan et al., 2020) and usually tends to develop 7–10 years earlier (Maas and Appelman, 2010; Piepoli, 2017). These reasons may partly explain the gender misbalance that was present in our study. In terms of age, a recent meta-analysis (Dibben et al., 2018) has reported the results of 47 studies performed in cardiac rehabilitations, suggesting the average age (58.3) in the study patients is close to the average age found in our study. Thus, the sample disequilibrium might be attributed to the reality of cardiac rehabilitations and their patients' sociodemographic profiles across different countries.

The present study has several strengths, such as the large sample size and the use of validated instruments. The study has also measured EC by employing a standardized computer-driven bicycle ergometer that allowed to objectively assess the cardiac function of the study patients and compare it with the MLHFQ results. In terms of research limitations, the study was completed in a single rehabilitation clinic; thus, the results are found in a selective individuals' group. Particularly, most of the study participants had mild-to-moderate HF, the significant majority were men, and all of them attended a single in-patient center, which may limit the generalization of our results. Thus, the findings should be interpreted with caution, when considering individuals with more advanced HF or individuals who do not attend rehabilitation programs after ACS.

Our study provided further knowledge of MLHFQ psychometric properties that may bolster the clinical knowledge of the application of this specific instrument in patients with CAD after ACS commonly presenting HF symptoms. This instrument may help to not only assess the subjectively perceived

HRQoL but also provide the benefit of patient-centered care to improve individual health outcomes and deliver more holistically appropriate interventions and customized medical care plans during cardiac rehabilitation. Also, culturally specific evaluation instruments can improve the communication between clinicians and patients that may result in more accurate symptom assessment and effective treatment. According to our results, Lithuanian clinicians may reconsider item 15 ("medical costs") during the assessment of HRQoL in cardiac rehabilitations, as it does not lead to any of the MLHFQ subscales and does not support the internal structure of this instrument.

CONCLUSION

This research offers additional details about the validation and psychometric properties of the MLHFQ in Lithuania. The MLHFQ is a reliable and valid instrument for measuring HRQoL in individuals with CAD following ACS. Our study also supports a three-factor solution for the MLHFQ that includes "physical," "social," and "emotional" dimensions. The study provides support for the MLHFQ potential use for future research and clinical practice in individuals undergoing cardiac rehabilitation with HF.

DATA AVAILABILITY STATEMENT

The study dataset is available upon reasonable request to NK, nijole.kazukauskiene@lsmuni.lt. The completed license agreement to use the MLHFQ can be provided upon request to the corresponding author. Due to the copyright regulations, the Lithuanian version of the MLHFQ can be provided upon the request to Laboratory of Behavioral Medicine, Neuroscience Institute, Lithuanian University of Health Sciences after the license agreement with University of Minnesota is completed.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Biomedical Research Ethics Committee for Biomedical Research at Lithuanian University of Health Sciences. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NK contributed to conceptualization, methodology, formal analysis, writing – review and editing, and investigation. JG-S contributed to writing – original draft, conceptualization, and investigation. JBu contributed to project administration, investigation, and writing – review and editing. JBr contributed to conceptualization and supervision. AB contributed to writing – review and editing. NK, JM, and VS contributed to supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This study of JBu was funded by the European Social Fund under the number 09.3.3-LMT-K-712-19-0127 "Development of Competencies of Scientists, other Researchers, and Students through Practical Research Activities".

ACKNOWLEDGMENTS

The authors are grateful for all the individuals who participated in this study and all research assistants, interviewers, computer, and laboratory technicians, including statistician E. Bovina, who conducted the statistical analysis.

REFERENCES

Amsterdam, E. A., Wenger, N. K., Brindis, R. G., Casey, D. E., Ganiats, T. G., Holmes, D. R., et al. (2014). AHA/ACC guideline for the management of patients with non–ST-elevation acute coronary syndromes: a report of the american college of cardiology/american heart association task force on practice guidelines. *J. Am. College Cardiol.* 64, e139–e228.

Barnett, S. D., Sarin, E. L., Henry, L., Halpin, L., Pritchard, G., and Speir, A. M. (2019). Confirmatory factor analysis of the Minnesota living with heart failure questionnaire among patients following open heart surgery for valve dysfunction. *Qual. Life Res.* 28, 267–275. doi: 10.1007/s11136-018-2022-1

Bennett, S. J., Oldridge, N. B., Eckert, G. J., Embree, J. L., Browning, S., Hou, N., et al. (2003). Comparison of quality of life measures in heart failure. *Nurs. Res.* 52, 207–216.

Bilbao, A., Escobar, A., Garcia-Perez, L., Navarro, G., and Quiros, R. (2016). The Minnesota living with heart failure questionnaire: comparison of different factor structures. *Health Qual. Life Outcom.* 14:23. doi: 10.1186/s12955-016-0425-7

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., and Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front. Public Health* 6:149. doi: 10.3389/fpubh.2018.00149

Brokalaki, H., Patelarou, E., Giakoumidakis, K., Kollia, Z., Fotos, N. V., Vivilaki, V., et al. (2015). Translation and validation of the Greek "Minnesota Living with Heart Failure" questionnaire. *Hellenic. J. Cardiol.* 56, 10–19.

Bussoni, M. F., Guirado, G. N., Roscani, M. G., Polegato, B. F., Matsubara, L. S., Bazan, S. G., et al. (2013). Diastolic function is associated with quality of life and exercise capacity in stable heart failure patients with reduced ejection fraction. *Braz. J. Med. Biol. Res.* 46, 803–808. doi: 10.1590/1414-431X20132902

Callus, E., Pagliuca, S., Bertoldo, E. G., Fiolo, V., Jackson, A. C., Boveri, S., et al. (2020). The monitoring of psychosocial factors during hospitalization before and after cardiac surgery until discharge from cardiac rehabilitation: a research protocol. *Front. Psychol.* 11:2202. doi: 10.3389/fpsyg.2020.02202

Carvalho, V. O., Guimarães, G. V., Carrara, D., Bacal, F., and Bocchi, E. A. (2009). Validação da versão em português do minnesota living with heart failure questionnaire. *Arquivos Brasileiros De Cardiologia* 93, 39–44. doi: 10.1590/s0066-782x2009000700008

Cordero, A., Rodriguez-Manero, M., Bertomeu-Gonzalez, V., Garcia-Acuna, J. M., Baluja, A., Agra-Bermejo, R., et al. (2021). New-onset heart failure after acute coronary syndrome in patients without heart failure or left ventricular dysfunction. *Revista Española de Cardiología* 74, 494–501. doi: 10.1016/j.rec.2020.03.011

Dibben, G. O., Dalal, H. M., Taylor, R. S., Doherty, P., Tang, L. H., and Hillsdon, M. (2018). Cardiac rehabilitation and physical activity: systematic review and meta-analysis. *Heart* 104, 1394–1402. doi: 10.1136/heartjnl-2017-312832

Erceg, P., Despotovic, N., Milosevic, D. P., Soldatovic, I., Zdravkovic, S., Tomic, S., et al. (2013). Health-related quality of life in elderly patients hospitalized with chronic heart failure. *Clin. Interv. Aging* 8, 1539–1546. doi: 10.2147/CIA.S53305

Falk, H., Ekman, I., Anderson, R., Fu, M., and Granger, B. (2013). Older patients' experiences of heart failure-an integrative literature review. *J. Nurs. Scholarsh.* 45, 247–255. doi: 10.1111/jnu.12025

Feeny, D. H., Eckstrom, E., Whitlock, E. P., and Perdue, L. A. (2013). *A primer for systematic reviewers on the measurement of functional status and health-related quality of life in older adults*. Rockville: Agency for Healthcare Research and Quality.

Fletcher, G. F., Ades, P. A., Kligfield, P., Arena, R., Balady, G. J., Bittner, V. A., et al. (2013). Exercise standards for testing and training: a scientific statement from the American Heart Association. *Circulation* 128, 873–934. doi: 10.1161/CIR.0b013e31829b5b44

Garin, O., Ferrer, M., Pont, A., Rue, M., Kotzeva, A., Wiklund, I., et al. (2009). Disease-specific health-related quality of life questionnaires for heart failure: a systematic review with meta-analyses. *Qual. Life Res.* 18, 71–85. doi: 10.1007/s11136-008-9416-4

Garin, O., Ferrer, M., Pont, A., Wiklund, I., Van Ganse, E., Vilagut, G., et al. (2013). Evidence on the global measurement model of the Minnesota Living with Heart Failure Questionnaire. *Qual. Life Res.* 22, 2675–2684. doi: 10.1007/s11136-013-0383-z

Garin, O., Herdman, M., Vilagut, G., Ferrer, M., Ribera, A., Rajmil, L., et al. (2014). Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures. *Heart Fail. Rev.* 19, 359–367. doi: 10.1007/s10741-013-9394-7

Garin, O., Soriano, N., Ribera, A., Ferrer, M., Pont, A., Alonso, J., et al. (2008). Validation of the spanish version of the minnesota living with heart failure questionnaire. *Rev. Esp. Cardiol.* 61, 251–259.

Gibbons, R. J., Balady, G. J., Bricker, J. T., Chaitman, B. R., Fletcher, G. F., Froelicher, V. F., et al. (2002). ACC/AHA 2002 guideline update for exercise testing: summary article. A report of the american college of cardiology/american heart association task force on practice guidelines (committee to update the 1997 exercise testing guidelines). *J. Am. Coll. Cardiol.* 40, 1531–1540.

Gonzalez-SaenzdeTejada, M., Bilbao, A., Ansola, L., Quiros, R., Garcia-Perez, L., Navarro, G., et al. (2019). Responsiveness and minimal clinically important difference of the Minnesota living with heart failure questionnaire. *Health Qual. Life Outcom.* 17:36. doi: 10.1186/s12955-019-1104-2

Haas, B. K. (1999). Clarification and integration of similar quality of life concepts. *Image J. Nurs. Scholarsh.* 31, 215–220. doi: 10.1111/j.1547-5069.1999.tb00483.x

Heo, S., Moser, D. K., Lennie, T. A., Zambroski, C. H., and Chung, M. L. (2007). A comparison of health-related quality of life between older adults with heart failure and healthy older adults. *Heart Lung.* 36, 16–24. doi: 10.1016/j.hrtlng.2006.06.003

Heo, S., Moser, D. K., Riegel, B., Hall, L. A., and Christman, N. (2005). Testing the psychometric properties of the minnesota living with heart failure questionnaire. *Nurs. Res.* 54, 265–272. doi: 10.1097/00006199-200507000-00009

Ho, C. C., Clochesy, J. M., Madigan, E., and Liu, C. C. (2007). Psychometric evaluation of the chinese version of the minnesota living with heart failure questionnaire. *Nurs. Res.* 56, 441–448. doi: 10.1097/01.NNR.0000299849.21935.c4

Iacoviello, M., and Antoncecchi, V. (2013). Heart failure in elderly: progress in clinical evaluation and therapeutic approach. *J. Geriatr. Cardiol.* 10, 165–177. doi: 10.3969/j.issn.1671-5411.2013.02.010

Journiac, J., Vioulac, C., Jacob, A., Escarnot, C., and Untas, A. (2020). What do we know about young adult cardiac patients' experience? a systematic review. *Front. Psychol.* 11:1119. doi: 10.3389/fpsyg.2020.01119

Kazukauskiene, N., and Burkauskas, J. (2018). Mental distress factors and exercise capacity in patients with coronary artery disease attending cardiac rehabilitation program. 25, 38–48. doi: 10.1007/s12529-017-9675-y

Kazukauskiene, N., Burkauskas, J., Macijauskienė, J., Mickuvienė, N., and Brožaitienė, J. (2019). Stressful life events are associated with health-related quality of life during cardiac rehabilitation and at 2-yr follow-up in patients with heart failure. *J. Cardiopulm. Rehabil. Prev.* 39, E5–E8. doi: 10.1097/HCR.0000000000000385

Khan, M. A., Hashim, M. J., Mustafa, H., Baniyas, M. Y., Al Suwaidi, S. K. B. M., AlKatheeri, R., et al. (2020). Global epidemiology of ischemic heart disease: Results from the global burden of disease study. *Cureus* 12:e9349. doi: 10.7759/cureus.9349

Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Rest. Dentis. Endodon.* 38, 52–54. doi: 10.5395/rde.2013.38.1.52

Lambrinou, E., Kalogirou, F., Lamnisos, D., Middleton, N., Sourtzi, P., Lemonidou, C., et al. (2013). Evaluation of the psychometric properties of the greek version of the minnesota living with heart failure questionnaire. *J. Cardiopul. Rehab. Preven.* 33, 229–233. doi: 10.1097/HCR.0b013e3182930cbb

Maas, A. H. E. M., and Appelman, Y. E. A. (2010). Gender differences in coronary heart disease. *Netherlands Heart J.* 18, 598–602. doi: 10.1007/s12471-010-0841-y

Middel, B., Bouma, J., de Jongste, M., van Sonderen, E., Niemeijer, M., and van den Heuvel, W. (2001). Psychometric properties of the Minnesota living with heart failure questionnaire (MLHF-Q). *Clin. Rehab.* 15, 489–500. doi: 10.1191/026921501680425216

Mogle, J., Buck, H., Zambroski, C., Alvaro, R., and Vellone, E. (2017). Cross-validation of the minnesota living with heart failure questionnaire. *J. Nurs. Scholarsh.* 49, 513–520. doi: 10.1111/jnu.12318

Moon, J. R., Jung, Y. Y., Jeon, E. S., Choi, J. O., Hwang, J. M., and Lee, S. C. (2012). Reliability and validity of the korean version of the minnesota living with heart failure questionnaire. *Heart Lung.* 41, 57–66. doi: 10.1016/j.hrtlng.2011.09.011

Morcillo, C., Aguado, O., Delas, J., and Rosell, F. (2007). Utility of the minnesota living with heart failure questionnaire for assessing quality of life in heart failure patients. *Rev. Esp. Cardiol.* 60, 1093–1096. doi: 10.1157/13111242

Mosterd, A., and Hoes, A. W. (2007). Clinical epidemiology of heart failure. *Heart* 93, 1137–1146.

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., et al. (2016). Executive summary: heart disease and stroke statistics–2016 update: a report from the american heart association. *Circulation* 133, 447–454. doi: 10.1161/CIR.0000000000000366

Munyombwe, T., Hofer, S., Fitzsimons, D., Thompson, D. R., Lane, D., Smith, K., et al. (2014). An evaluation of the minnesota living with heart failure questionnaire using rasch analysis. *Qual. Life Res.* 23, 1753–1765. doi: 10.1007/s11136-013-0617-0

Napier, R., McNulty, S. E., Eton, D. T., Redfield, M. M., AbouEzzeddine, O., and Dunlay, S. M. (2018). Comparing measures to assess health-related quality of life in heart failure with preserved ejection fraction. *JACC Heart Fail.* 6, 552–560. doi: 10.1016/j.jchf.2018.02.006

O'Gara, P. T., Kushner, F. G., Ascheim, D. D., Casey, D. E. Jr., Chung, M. K., de Lemos, J. A., et al. (2013). 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: executive summary: a report of the american college of cardiology foundation/american heart association task force on practice guidelines: developed in collaboration with the american college of emergency physicians and society for cardiovascular angiography and interventions. *Catheter. Cardiovasc. Interv.* 82, E1–E27. doi: 10.1002/ccd.24776

Parajon, T., Lupon, J., Gonzalez, B., Urrutia, A., Altimir, S., Coll, R., et al. (2004). Use of the minnesota living with heart failure quality of life questionnaire in spain. *Rev. Esp. Cardiol.* 57, 155–160. doi: 10.1016/s1885-5857(06)60104-7

Piepoli, M. F. (2017). *2016 European Guidelines on cardiovascular disease prevention in clinical practice*. New York, NY: Springer.

Piepoli, M. F., Corra, U., Benzer, W., Bjarnason-Wehrens, B., Dendale, P., Gaita, D., et al. (2010). Secondary prevention through cardiac rehabilitation: from knowledge to implementation. a position paper from the cardiac rehabilitation section of the european association of cardiovascular prevention and rehabilitation. *Eur. J. Cardiovasc. Prev. Rehabil.* 17, 1–17. doi: 10.1097/HJR.0b013e3283313592

Qureshi, W. T., Zhang, Z.-M., Chang, P. P., Rosamond, W. D., Kitzman, D. W., Wagenknecht, L. E., et al. (2018). Silent myocardial infarction and long-term risk of heart failure: the ARIC study. *J. Am. College Cardiol.* 71, 1–8. doi: 10.1016/j.jacc.2017.10.071

Rajati, F., Feizi, A., Tavakol, K., Mostafavi, F., Sadeghi, M., and Sharifirad, G. (2016). Comparative evaluation of health-related quality of life questionnaires in patients with heart failure undergoing cardiac rehabilitation: a psychometric study. *Arch. Phys. Med. Rehabil.* 97, 1953–1962. doi: 10.1016/j.apmr.2016.05.010

Rector, T. S., and Cohn, J. N. (1992). Assessment of patient outcome with the Minnesota Living with Heart Failure questionnaire: reliability and validity during a randomized, double-blind, placebo-controlled trial of pimobendan. pimobendan multicenter research group. *Am. Heart J.* 124, 1017–1025. doi: 10.1016/0002-8703(92)90986-6

Sneed, N. V., Paul, S., Michel, Y., VanBakel, A., and Hendrix, G. (2001). Evaluation of 3 quality of life measurement tools in patients with chronic heart failure. *Heart Lung.* 30, 332–340. doi: 10.1067/mhl.2001.118303

Staniute, M., Brozaitiene, J., Burkauskas, J., Kazukauskiene, N., Mickuviene, N., and Bunevicius, R. (2015). Type D personality, mental distress, social support and health-related quality of life in coronary artery disease patients with heart failure: a longitudinal observational study. *Health Qual. Life Outcomes* 13:1. doi: 10.1186/s12955-014-0204-2

Ware, J. E., and Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med. Care* 30, 473–483. doi: 10.1097/00005650-199206000-00002

Zahwe, M., Isma'eel, H., Skouri, H., Al-Hajje, A., Rachidi, S., Tamim, H., et al. (2020). Validation of the arabic version of the minnesota living with heart failure questionnaire. *Heart Lung.* 49, 36–41. doi: 10.1016/j.hrtlng.2019.10.006

Conflict of Interest: JG-S served as a consultant at FACITtrans. In the past several years JBu worked as a consultant to Cogstate, Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 14 February 2022
doi: 10.3389/fpsyg.2022.804885

The Measurement Properties and Acceptability of a New Parent–Infant Bonding Tool ('Me and My Baby') for Use in United Kingdom Universal Healthcare Settings: A Psychometric, Cross-Sectional Study

Tracey Bywater[1], Abigail Dunn[2], Charlotte Endacott[3], Karen Smith[4], Paul A. Tiffin[1], Matthew Price[5] and Sarah Blower[1]*

[1] Department of Health Sciences, Hull York Medical School, University of York, York, United Kingdom, [2] Department of Social Policy and Social Work, University of York, York, United Kingdom, [3] Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, United Kingdom, [4] Rotherham Doncaster and South Humber NHS Foundation Trust, Doncaster, United Kingdom, [5] Little Minds Matter Bradford Infant Mental Health Service, Bradford District Care NHS Foundation Trust, Bradford, United Kingdom

OPEN ACCESS

Edited by:
Paola Gremigni,
University of Bologna, Italy

Reviewed by:
María De La Fe Rodriguez,
Universidad Nacional de Educación a Distancia, Spain
Mike Horton,
University of Leeds, United Kingdom

****Correspondence:***
Tracey Bywater
tracey.bywater@york.ac.uk

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *29 October 2021*
Accepted: *03 January 2022*
Published: *14 February 2022*

Citation:
Bywater T, Dunn A, Endacott C, Smith K, Tiffin PA, Price M and Blower S (2022) The Measurement Properties and Acceptability of a New Parent–Infant Bonding Tool ('Me and My Baby') for Use in United Kingdom Universal Healthcare Settings: A Psychometric, Cross-Sectional Study. Front. Psychol. 13:804885. doi: 10.3389/fpsyg.2022.804885

Introduction: The National Institute for Health and Care Excellence (NICE) guidelines acknowledge the importance of the parent–infant relationship for child development but highlight the need for further research to establish reliable tools for assessment, particularly for parents of children under 1 year. This study explores the acceptability and psychometric properties of a co-developed tool, 'Me and My Baby' (MaMB).

Study design: A cross-sectional design was applied. The MaMB was administered universally (in two sites) with mothers during routine 6–8-week Health Visitor contacts. The sample comprised 467 mothers (434 MaMB completers and 33 'non-completers'). Dimensionality of instrument responses were evaluated via exploratory and confirmatory ordinal factor analyses. Item response modeling was conducted via a Rasch calibration to evaluate how the tool conformed to principles of 'fundamental measurement'. Tool acceptability was evaluated via completion rates and comparing 'completers' and 'non-completers' demographic differences on age, parity, ethnicity, and English as an additional language. Free-text comments were summarized. Data sharing agreements and data management were compliant with the General Data Protection Regulation, and University of York data management policies.

Results: High completion rates suggested the MaMB was acceptable. Psychometric analyses showed the response data to be an excellent fit to a unidimensional confirmatory factor analytic model. All items loaded statistically significantly and substantially (>0.4) on a single underlying factor (latent variable). The item response modeling showed that most MaMB items fitted the Rasch model. (Rasch) item reliability was high (0.94) yet the test yielded little information on each respondent, as highlighted by the relatively low 'person separation index' of 0.1.

Conclusion and next steps: MaMB reliably measures a single construct, likely to be infant bonding. However, further validation work is needed, preferably with 'enriched population samples' to include higher-need/risk families. The MaMB tool may benefit from reduced response categories (from four to three) and some modest item wording amendments. Following further validation and reliability appraisal the MaMB may ultimately be used with fathers/other primary caregivers and be potentially useful in research, universal health settings as part of a referral pathway, and clinical practice, to identify dyads in need of additional support/interventions.

Keywords: parent, baby, measure, psychometrics, bonding

INTRODUCTION

As mothers are typically primary caregivers, the current study evaluated the Me and My Baby (MaMB) for use by mothers. Maternal bonding can be defined as a mother's emotional connection and feeling toward her child (Condon, 1993). Bonding is often conflated with attachment. Whilst the constructs are related, they are distinct (Bowlby, 1982; Redshaw and Martin, 2013). Maternal bonding refers to a mother's (typically self-reported) emotional connection and feelings toward their child. Attachment on the other hand, refers to an infant's expectations of their caregiver's responses and the pattern of their own behavior, e.g., when activated in response to a perceived threat. Attachment typically develops from 6 months, whereas a mother's bond to the infant begins to develop during pregnancy. Stronger bonding is theoretically linked to more frequent expression of behaviors such as maternal sensitivity and emotional availability (Feldman et al., 1999), which in turn foster positive interactions within the dyad and promote social and emotional development, including the development of secure attachment in the infant (Ainsworth et al., 1978; Le Bas et al., 2019).

Two systematic reviews (Branjerdporn et al., 2017; Le Bas et al., 2019) indicate that strong maternal bonding in pregnancy is associated with optimal child developmental outcomes. The Le Bas et al. (2019) review also suggested that higher affective postnatal parent-infant bond was predictive of positive child development outcomes. Both reviews suggested the findings should be interpreted with caution due to the relative paucity of studies in this area and highlighted the need for more robust self-report measures of bonding.

There are currently no agreed, standardized, methods for identifying mother/parent–infant dyads who may benefit from additional support around bonding and relationships in England. Although Health Visitors (HVs) work directly with parents some research suggests that they may struggle to consistently identify problems in the parent–infant relationship (Wilson et al., 2010; Appleton et al., 2013; Kristensen et al., 2017; Elmer et al., 2019). Relevant the National Institute for Health and Care Excellence (NICE) guidelines acknowledge the importance of parent–infant relationship for child development and parent mental health but highlight the need for further research to establish reliable tools for assessment, particularly for parents of children under the age of 1 year (NICE, 2012; NICE., 2015).

There is a distinct need for validated, robust measures to be administered universally to identify and support families who may struggle with their parent-infant relationship. Parent–infant relationship is a key focus in the Early Years High Impact Area 2: supporting good parental mental health (Public Health England [PHE], 2020) due to the risks to subsequent child social and emotional development arising from poor parent–infant relationships (Fearon et al., 2010; Cassidy et al., 2013). A reliable, valid, identification tool could allow services to more confidently signpost parents who may benefit to one of the emerging evidence-based interventions (Barlow et al., 2010, 2016; Wright et al., 2015; Facompre et al., 2018).

A very limited number of brief parent self-report tools exist that assess maternal-infant bonding, are freely available, and have some reliability and validity (Kane, 2017; Blower et al., 2019; Gridley et al., 2019; Wittkowski et al., 2020), for example; Maternal Attachment Inventory (MAI; Müller, 1994); Maternal Postnatal Attachment Scale (MPAS) (Condon and Corkindale, 1998); Postpartum Bonding Questionnaire (PBQ) (Brockington et al., 2006); Mother Infant Bonding Scale (MIBS) (Taylor et al., 2005). However, most are not widely used, or have been validated with a small sample (for further discussion see Le Bas et al., 2019; Wittkowski et al., 2020). A further two reviews, Blower et al. (2019) and Gridley et al. (2019) were undertaken to explore which measures would be acceptable, reliable, and valid for a large randomized controlled trial of a parenting intervention for parents of infants and toddlers and it was found that choice of measures was very limited (the trial was led by TB, the first author. For the protocol see Bywater et al., 2018).

The 19-item MPAS, which has preliminary evidence of reliability and validity (Kane, 2017; Wittkowski et al., 2020) is the most used tool when linking maternal-infant bonding to later child development outcomes (Le Bas et al., 2019). The MPAS was piloted (with the involvement of the first and second authors) with 347 mothers in universal health visiting services (Bird et al., 2021; Dunn et al., 2021) as part of Better Start Bradford – a 10-year National Lottery Community Fund project aimed at improving the socio-emotional development, nutrition and communication skills of children aged 0–3 living in deprived multi-ethnic communities (Dickerson et al., 2016). The pilot concluded that the MPAS could not be recommended for use in health visiting services in Bradford to assess parent-infant relationship due to; little variation in the responses of the 225 who completed the MPAS in

English; an unexpected ceiling effect; issues with scoring, parental acceptability and understanding. The E-SEE trial found similar findings, with lack of variation in scores on a sample of 341 (Bywater et al., 2018).

Using the learning from the MPAS pilot the study team co-developed a new tool, MaMB, in an iterative process via workshops and interviews with Health Visitors, Clinical Psychologists, service staff, Managers and parental input, to address the issues highlighted in the MPAS pilot. Prior to a measure being recommended for use in any context, evidence of the measurement properties should be established (Cooper, 2019). Psychometric properties comprise two overarching dimensions – validity and reliability. Validity is defined as the degree to which an instrument measures the construct(s) it purports to measure, and reliability is the degree to which a measure is free from measurement error (de Vet et al., 2015). Acceptable reliability is thus a necessary, though not sufficient, condition for achieving valid scores from an instrument. ‘Reliability’ also relates to the important concept of ‘test information’; that is, the trait level at which the instrument is most capable of discriminating between test takers/respondents. Thus, a test’s ‘information curve’ has important implications for how it is optimally used in practice; for example, when identifying a screening cut-off score.

This study was therefore intended to evaluate the measurement model for the MaMB and acceptability when implemented in routine practice, as a prerequisite to further studies aiming to establish validity of the tool. The main aim was to address previous paucity and quality of available tools to assess parent (mother)–infant relationship, specifically bonding, by developing a measure for use in research as well as universal health settings as part of a referral pathway, and potentially clinical practice, to identify dyads in need of additional support or interventions. The research objectives for this study were:

(1) To explore MaMB pilot data to determine the item and test properties in relation to dimensionality and reliability, in terms of both internal consistency and test information; and
(2) To identify any necessary revisions to MaMB following the results of our psychometric analysis.

These findings would have implications for which items would be retained in a final version of instrument, and how the scores might be best summarized and used in practice. The work also paves the way for validation studies.

MATERIALS AND METHODS

The Tool Under Investigation

The MaMB questionnaire (for further information see **Supplementary Measure 1**, and the protocol at https://osf.io/q3hmf/) has 11 items presented in a user-friendly format. Responses are indicated using a four-point Likert scale (‘never,’ ‘sometimes,’ ‘often,’ or ‘always,’ scored 0–3 with four reversed scored items). The language of items is simple to understand with a reading age of approximately 12, similar to that for popular magazines. A free text box is also included to give mothers the opportunity to record any comments or concerns they have about their relationship with their infant. Lower scores indicate a stronger affective bond.

Research Questions

RQ1: Is the MaMB acceptable to mothers of infants (aged 6–8 weeks) and HVs when administered in a universal healthcare setting?

(a) As a proportion of all eligible dyads, how many complete the MaMB?
(b) What are the reasons given for non-completion?
(c) Are the free text boxes completed by parents and what information is being recorded/reported in them?

RQ2: What are the measurement properties of the MaMB?

(a) What is the most plausible dimensionality (factor structure) of the MaMB?
(b) Does the scale (or subscales if applicable) of the MaMB demonstrate acceptable levels of internal consistency?
(c) According to item response modeling, do the items demonstrate an acceptable fit to the Rasch model, implying that the summed scores from the instrument can be used as a ‘sufficient summary statistic’?
(d) What is the relative level of information yielded for respondents by the test (or putative scales), and where might a potential cut-off score be best placed that most accurately differentiates between two groups of test-takers?

Design

A cross-sectional design was applied.

A briefing was prepared in partnership with Rotherham Doncaster and South Humber NHS Foundation Trust (RDaSH) to support the training of HVs in the use of the tool. The briefing covered the purpose of the tool, how to introduce it to families, how to score it and how to interpret the scores.

The MaMB was implemented universally (in two RDaSH localities) with eligible mothers during the 6–8 weeks routine HV contact following completion of the core mandated elements of the visit.

Health visitors asked mothers to complete a paper version of the tool, with support if needed or requested. During tool completion HVs were expected to use their professional skills to discuss with parents their relationship with their infant. If HVs were unable to complete the tool (e.g., due to time constraints) they would record the reason(s) for non-completion.

Health visitors inputted the responses electronically into the case management software (SystmOne) co-developed template to include; if tool administration was attempted, and if not why, and if tool administration had been abandoned prior to completion. The template also captured responses to all 11 items, and the free text responses to the open question on the back page of the paper tool, and HVs comments on the interaction. Key demographic variables were also recorded to adequately describe the sample’s characteristics and to support subgroup analyses.

The research team received anonymized (numerical and free text) data extracted from SystmOne, and a small number of key demographic characteristics such as age, ethnicity, and parity.

Study Setting

Two RDaSH sites in Northern England implemented the MaMB at the 6–8 weeks universally mandated HV contact.

Inclusion/Exclusion Criteria

All mothers of a child aged 6–8 weeks living in the sites were eligible for the study.

If a parent had opted out of NHS digital they may have completed the MaMB but were not included in the study (in England, NHS patients can choose to opt out of their confidential patient information being used for research and planning).

Consent

This study received ethical approval on 21st August 2020 by South Central - Berkshire B Research Ethics Committee, United Kingdom, Ref: 20/SC/0266, Integrated Research Application System (IRAS) 201, project ID: 273708.

Parents were given a MaMB Participant Information Sheet (V2.0 17th August 2020; see **Supplementary Information Sheet 1**) at a visit prior to the 6–8-week check to give them time to read and understand why they will be asked to complete the MaMB.

Written consent from mothers completing the MaMB, and for the non-identifiable fully anonymized, data to be shared with the research team, was not required. This was because:

(1) The research team only accessed anonymized data. Data were restricted to the minimum needed to describe the sample and to conduct the proposed analyses of measurement properties and acceptability. Free text boxes, where completed, and were screened by an authorized RDaSH employee to remove any identifiable information prior to data sharing.
(2) There was no risk of harm to participants from completing the MaMB. The tool was one of several used by HVs to conduct a broad needs assessment, as is standard at the 6–8-week contact. The MaMB supplemented existing tools and was implemented in addition to standard care. HVs are trained and well equipped to support mothers who may be struggling to bond with their baby.
(3) It was deemed essential that the MaMB sample was representative of mothers of young infants in the research site so that the study findings are generalizable. Introducing an informed consent process would likely have led to selection bias, arising from parent and practitioner characteristics and attitudes.
(4) There is a clear value and benefit from doing the research, i.e., a need for a short, easy-to-administer, valid and reliable measure to support practitioners to identify families experiencing difficulties in their parent-infant relationship. The MaMB has been co-developed by academics, psychologists and HVs with parental input to address this gap, it is vital that this measure is tested before it can be recommended for use more widely.

Sample Size

The average number of live births per year in the year prior to the study was 3460 in Site 1 (Doncaster) and 3000 in Site 2 (North Lincolnshire), which would yield approximately 538 births per month. Assuming a conservative 50% completion rate (allowing for potential implementation/uptake barriers such as time constraints, parent refusal or practitioner non-compliance, time lag in implementation and data entry) we anticipated 269 MaMBs would be completed per month. To construct a sample large enough to support the analysis of psychometric properties we proposed a sample of 673 over a 10-week period. Based on a 50% completion rate, the overall sample would include a further 673 non-completers to explore acceptability (total $n = 1346$). Please note this sample size was calculated pre-COVID-19.

Psychometric Analyses

RQ1

To assess acceptability of the tool reported the proportion of participants who were recorded as being offered the tool but either refused, or failed to complete, it. Where data were available descriptive analysis of the reasons for refusal was to be produced.

Key demographic characteristics (age, parity, ethnicity, English as an additional language) of completers and non-completers were to be presented in contingency tables as either frequency counts or means for descriptive purposes.

A frequency count was intended to determine the proportion of completers who used the free-text box to expand on their answers. Free-text comments were to be summarized in a brief narrative.

RQ2

Dimensionality and Internal Consistency Reliability

The sample was originally intended to be randomized into exploratory and confirmatory ('validation') datasets, if the data obtained were sufficient to support this approach. Initially dimensionality was planned to be explored in the former data subset using parallel analysis (see below for details) (Horn, 1965). Once this had been established, it would be followed by an exploratory factor analysis (EFA) of exploratory portion of the response data. The potential factor structures elicited would then be tested using confirmatory factor analyses (CFA) on the confirmatory (validation) dataset. Internal reliability consistency of the postulated subscales would then be examined. The findings of these analyses were intended to indicate whether it is appropriate to summarize bonding via several subscales or simply by a single total overall score for the MaMB.

The parallel analysis would be performed using unweighted least squares (ULS) as the estimation method (Horn, 1965; Lorenzo-Seva and Ferrando, 2006). In a parallel analysis the maximum plausible number of factors to be retained is indicated at the point where the eigenvalues of the randomly generated data exceed those of the actual data. A series of EFAs was expected to be then performed to aid interpretation of any factors underlying the response patterns observed. Oblique (geomin) rotations were to be used in the factor analyses, assuming that, as in almost all psychological measures, underlying latent traits would be correlated with each other to some extent. The EFAs will be repeated, again using a geomin rotation, to derive standard errors (and thus standardized Z scores) for the factor loadings to evaluate their relative statistical significance (Asparouhov and Muthén, 2009). All EFAs and CFAs were to be conducted in Mplus version 6.1 employing robust weighted least squares (WLSMV) as the estimation method, or 'full information maximum likelihood,' as appropriate.

Internal reliability consistency for the putative subscales based on the CFA structure was to be evaluated using Cronbach's alpha and McDonald's omega. Cronbach's alpha may be a poor index of internal reliability where tau-equivalence (equality of factor loadings across items in a scale) does not hold (Raykov, 1997). In this respect McDonald's omega is reported to represent a more accurate estimate of the extent to which items in a scale measure a unidimensional underlying construct.

Item Response Modeling

Item response modeling and theory (IRT) is based on the modified factor analysis of binary and categorical data. Within the family of IRT models Rasch analysis was originally developed for the exploration of dichotomous responses to test items (Rasch, 1960), though was subsequently extended to accommodate polytomous data. Rasch analysis can be used to create interval metrics of both item difficulty and respondent ability from ordinal (ordered categorical) or binary (dichotomous) response data. The Rasch model assumes that all items are identical in terms of their ability to discriminate between respondents according to ability/trait (i.e., equality of item factor loadings in classical factor analytic terms). For the present Rasch analysis the software package Winsteps version 4.01 was used (Linacre, 2017). A partial credit model was applied to the categorical MaMB item responses. In a Rasch analysis reliability can be appraised in several ways. Specifically, the person reliability coefficient relates to the replicability of the ranking of abilities while the person separation index represents the signal to noise ratio and estimates the ability of a test to reliably differentiate different levels of ability within a cohort (Wright and Masters, 1982).

Power issues in Rasch analysis are a matter for debate with some authors suggesting that around 200 respondents are required to accurately estimate item difficulty whilst others suggest as few as 30 participants may be required in well-targeted tests (i.e., those where difficulty is well matched to ability) (Goldman and Raju, 1986; Linacre, 1994; Baur and Lukes, 2009). Thus, this study should be adequately powered to estimate item properties from both Rasch analysis as well as the factor analyses, the latter of which could be considered re-parameterized two parameter logistic regression IRT models. Thus, the fit of items to the Rasch model was to be assessed and any potential sources of misfit diagnosed. This will be important in deciding whether it is appropriate to summarize the scores on the scale/s as summed totals. Moreover, the Rasch calibration was intended to allow the evaluation of test information, which would indicate to what extent the test is able to differentiate test-takers across the putative trait levels under evaluation (assumed to be 'perceived bonding with baby').

Data Handling and Sharing

Fully anonymized data was exported from SystmOne and shared with the study team via the University of York secure drop off service, which securely encrypts data. Data management is compliant with the General Data Protection Regulation (GDPR) and University of York data management policies. The custodian of data, Professor Tracey Bywater (Chief Investigator), is the contact point for any data management queries.

RESULTS

The pilot ran 10th September 2020 to 1st December 2020, and the MaMB was administered either face to face or over the telephone depending on COVID-19 restrictions at the time of administration.

See **Figure 1** for a flow of participants through the study.

The 434 response rate from the eligible 928 women equates to a 47% response rate, close to the predicted 50%.

The target sample size of 673 for MaMB completion was not achieved, and we only have data for 33/494 women who did not complete the MaMB rather than the proposed 673. The birth rate was lower than expected, and HVs changed to telephone rather than face to face visits during the study due to COVID.

Results will be presented in order of the research questions.

> RQ1: Is the MaMB acceptable to mothers of infants (aged 6–8 weeks) and HVs when administered in a universal healthcare setting?

(a) As a proportion of all eligible dyads, how many complete the MaMB?
(b) What are the reasons given for non-completion?
(c) Are the free text boxes completed by parents and what information is being recorded/reported in them?

Table 1 shows the characteristics of the sample who completed the MaMB. The sample appears to represent the local population regarding ethnicity (83% white British, 10% White other, 7% Black, Asian, Multi-ethnic and other) and language (80% English as a first language, 6% missing). Although the numbers are small and we cannot draw conclusions from them, the 33 non-completers appeared to differ on ethnicity and language, which may be a reason for not completing the MaMB, e.g., 24% were white 'other' in the non-completers, compared to 10% in the completers. Likewise, 38% of non-completers needed an interpreter compared to 14% from the completers. Although 461 cover sheets for non-completers were missing, there was minimal missing data at item-level for those that were returned.

From the 434 respondents who completed a MaMB 50 had one or more missing items. Scores from the 384 who fully completed the MaMB tool suggest that the sample had positive relationships with their baby, mean = 1.2 (SD 1.6), with a median summed score of 1 (inter-quartile range 0–2) from a possible 33 (the

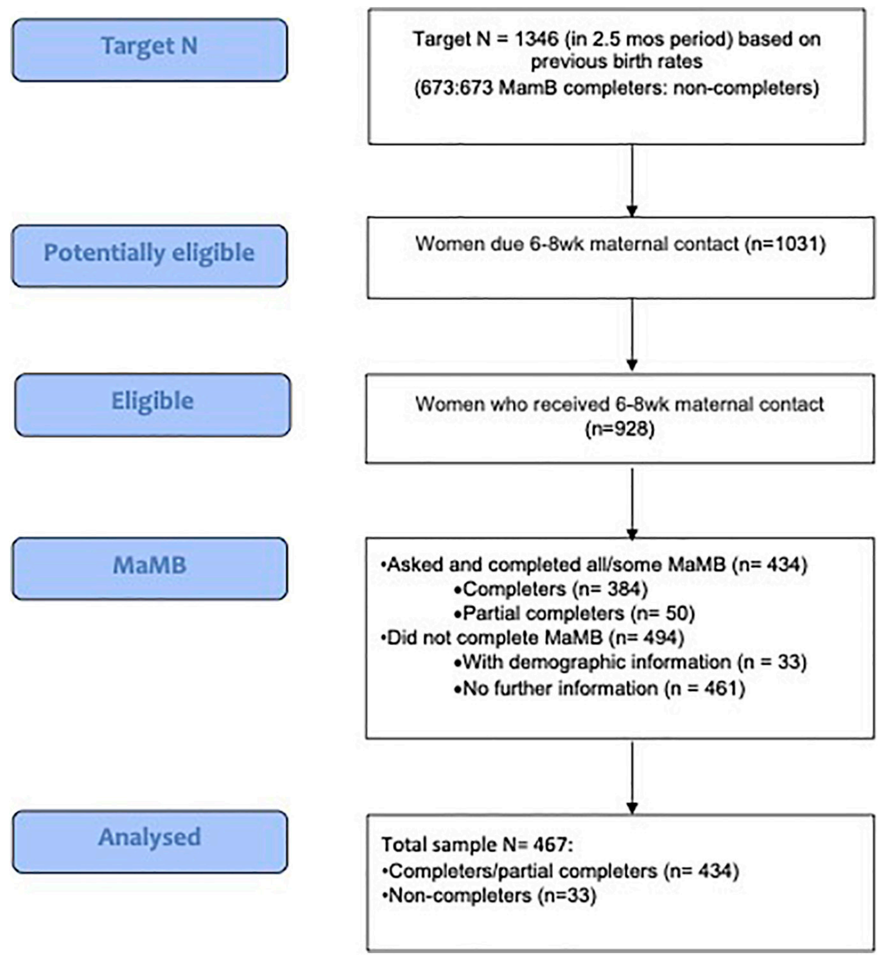

FIGURE 1 | Flow chart of participants.

lower the score indicating the more positive the perception of the mother-baby relationship), and a range of 0–15.

Twenty-nine respondents (parents and HVs) completed the free-text box with some mothers saying they felt guilty that they could not give more time to their baby or felt less than positively to toward their child at times, e.g;

> *"I feel guilty for having less positive feelings especially when he is screaming"*
> *"I feel I need time by myself sometimes, but feel guilty that I feel like that as a mum"*

Four mothers mentioned that they had not been separated from their baby yet, so items 8 and 10 were not applicable.

RQ2: *What are the measurement properties of the MaMB?*

From 467 mothers 33 had no MaMB questionnaire data whatsoever, leaving 434 participants with some response data. The original plan was to divide up the data, randomly, into a training and validation set (see section "Materials and Methods"). However, due to lack of variance in some of the item responses this was not possible. That is, dividing the dataset into two portions created items where little or no variation in responses were observed in some cases, rendering estimation of factor models impossible. Therefore, the entire dataset was explored in relation to its dimensionality.

Dimensionality

Firstly, a parallel analysis was conducted using the software FACTOR. This generates pseudorandom data, with the same dimensions as the real data. This process was adapted for use with the ordinal response data using polychoric matrices. Missing data values were handled using hot deck multiple imputation (Lorenzo-Seva and Van Ginkel, 2016). The results of the parallel analysis are shown in **Table 2**. These clearly indicate that there is a maximum of one factor (latent variable) underlying the response structure. This is evidenced clearly by the fact that the first latent variable explains around 60% of the variance in the indicators (item responses). However, a second postulated latent variable explains less variance than that found in a second latent variable for the pseudorandom data. The reliability, as indexed by Cronbach's alpha was 0.64 (standardized Cronbach's alpha 0.92) and McDonald's Omega value of 0.92.

TABLE 1 | Characteristics of completers (*N* = 434) and non-completers (*N* = 33).

	Completers (*N* = 434)		Non-completers (*N* = 33)	
	Count	**Percent**	**Count**	**Percent**
Site				
Doncaster (Site 1)	256	59%	21	64%
North Lincolnshire (Site 2)	178	41%	12	36%
Mother age (in years)				
Mean (*SD*)	28.45 (5.76)	/	29.25 (5.17)	/
Min	16	/	21	/
Max	43	/	43	/
Child age (in weeks)				
Mean (*SD*)	6.69 (1.69)	/	8	/
Min	4	/	6	/
Max	25	/	31	/
Ethnicity				
White British	359	83%	16	49%
White Other	43	10%	8	24%
Asian/Asian British	13	3%	0	0
Black African/Caribbean/Black British	5	1%	3	9%
Mixed/Multi-ethnic	2	0.5%	1	3%
Other	9	2%	1	3%
Missing	3	0.5%	4	12%
Mother's first language is English				
Yes	348	80%	15	46%
No	59	14%	13	39%
Missing	27	6%	5	15%
Interpreter needed (for non-first language English)				
Yes	8	14%	5	38%
No	50	85%	7	54%
Missing	1	1%	1	8%
First child				
Yes	195	45%	9	27%
No	235	54%	20	61%
Missing	4	1%	4	12%

Table includes a descriptive summary of available data from the 33 women who did not complete a MaMB but their health visitor completed a cover sheet.

The goodness of fit index for the one factor EFA was 0.985 (95% confidence intervals, derived via bootstrapping, 0.985–0.989). The psychometric properties of the items are shown in **Table 3**. For the standardized covariance matrix (polychoric correlations) as estimated from an ordinal factor analysis of the items of the MaMB scale, using the FACTOR software package see **Supplementary Table 1** provided.

This unidimensional structure was confirmed by examining the fit to a single factor confirmatory factor analytic model within the Mplus v8.4 software environment. This confirmatory factor analysis (CFA) was adapted for the ordinal nature of the response data, using robust weighted least squares as the estimation method (WLSMV). There were technical difficulties estimating a one factor model due to the low variance in items 4 and 5 and their collinearity with responses to items 10 and 11 respectively (that is, responses to the latter items were almost wholly associated with response to the former). Specifically, the correlation between item 4 ('difficult') and item 10 ('apart') was 0.987. That between item 5 ('need') and item 11 ('play') was also 0.987. Consequently, items 4 and 5 (which exhibited the lowest variance of the pairs were dropped from the CFA). When the CFA was repeated with the remaining nine items the one factor model showed a good fit to the data; the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) fit indices were 0.94 and 0.92 respectively ($\geq$0.90 usually is taken as acceptable fit, whilst values over 0.95 indicate good fit). Combining positive and negative worded items in a single scale can sometimes artificially lead to method effects. That is, these item types can sometimes show dependency on each other that manifest as correlated model residuals or 'artifactors' (Marsh, 1996). For this reason the residuals from negatively worded items were permitted to correlate within the CFA model to evaluate if this resulted in improved fit. However, this was not the case, with fit, if anything, deteriorating slightly (the TLI reduced from 0.92 to 0.91). Moreover, the modification indices did not suggest that fit would be significantly improved by permitting correlated residuals between items. The issue of dependency between items was also evaluated as part of the Rasch calibration (see below).

TABLE 2 | Results from a parallel analysis, adapted for ordinal data.

Factor	Variable Real-data % of variance	Mean of random % of variance	95th percentile of random % of variance
1st	61.4*	18.4	21.9
2nd	10.1	16.2	18.6
3rd	6.7	14.3	16.1
4th	5.8	12.6	14.2
5th	5.0	10.7	12.3
6th	3.1	9.0	10.5
7th	2.6	7.3	9.0
8th	2.3	5.5	7.4
9th	1.9	3.9	5.8
10th	1.2	2.1	4.0
11th	0.0	0.0	0.0

**Note only the percentage of variance explained by the first factor exceeds that observed for the random data.*

The factor loadings demonstrate a substantial (>0.4), positive and significant ($p < 0.01$) magnitude of loadings for all nine MaMB items included. Negative items were reverse coded so that the latent variable and the item factor loadings were interpretable. Having established the unidimensional structure of the data it appeared appropriate to progress to a Rasch calibration of the MaMB items.

Rasch Analysis

The Rasch calibration results yielded much useful diagnostic information on the MaMB questionnaire. As highlighted earlier the scale reliability itself was moderate to high. Indeed, the item reliability estimated by the Rasch calibration was 0.94. However, the person separation index (which include 'extreme' and 'non-extreme' persons) was only 0.10. The person separation index reflects the number of groups that can be plausibly differentiated by the scale with acceptable precision. It represents a signal to noise ratio in the scale. Thus, the MaMB scale had virtually no ability to differentiate respondents. This was no doubt a reflection on the lack of observed variance in responses in the study sample. Nevertheless, in terms of scale development and future research it is useful to explore the item 'difficulties' (or 'endorsibility' in this case), as well as the fit statistics. These are shown below in **Table 4**. The z standardized fit, along with the difficulty/endorsibility and standard error (reflected in the diameter of each bubble) are also shown in the 'bubble plots' in **Figures 2**, **3**. In the Rasch context 'fit' in this sense refers to which the item responses follow a Guttman sequence (Rasch, 1960). That is, as the ability or trait increases the respondent or test-taker tends to be observed to give a higher scoring category of response, allowing for the play of chance, e.g., 00101011122212212222233233333. Items where responses are too predictable 'overfit' the model. Those that are more erratic are described as 'underfitting.' The former tends to indicate redundant items, that may be dependent on responses to other items. Underfitting items can distort or degrade the measurement properties of the scale. 'Infit' refers to fit where an item 'difficulty' is well matched to the level of trait or ability in a test taker. That is, for example, for a right/wrong maths question the person who is well matched would have a 50:50 chance of either a correct or incorrect answer. In this case 'well targeted' items would tend to show a reasonable spread of responses for a set of test takers with trait levels that are matched to the item endorsibility. Conversely, 'outfit' refers to fit (conformity to the Rasch model) where item difficulty is not well matched to the test taker's trait or ability level. These distinctions between infit and outfit tend to be more pertinent to knowledge tests, than trait assessments, however. As can be seen from **Table 4** and **Figures 2**, **3** overall, the MaMB items tend to conform reasonably well to the Rasch model. However, there are four key issues.

(1) The items seem very easy (or in the case of negatively worded items- very hard) to endorse. This can be seen by the 'measure' estimates that tend to be around or above the zero point- a standardized trait (estimate) derived from the item responses.
(2) A couple of items tend to overfit the model: 'enjoy' (item 1) and 'irritated' (item 2). These tend to be somewhat overly predictable from the responses to the other items. However, this observation should be viewed cautiously as only the z standardized fit showed misfit,

TABLE 3 | Psychometric properties of the MAMB items, including exploratory factor analysis results, assuming one underlying factor (dimension).

MaMB item (abbreviated wording)	Item mean (*SD*)	Item-total correlation	Cronbach's alpha with item removed*	Factor loading	Communality
(1) Enjoy looking after baby	0.08 (0.29)	0.64	0.58	0.849	0.721
(2) Feel irritated with baby	0.08 (0.27)	0.52	0.60	0.709	0.502
(3) Affectionate toward baby	0.04 (0.2)	0.51	0.60	0.835	0.698
(4) Feel baby is being difficult	0.02 (0.15)	0.33	0.63	0.675	0.456
(5) Can work out baby's needs	0.54 (0.58)	0.57	0.66	0.489	0.239
(6) Can't do enjoyable things because of baby	0.21 (0.43)	0.55	0.61	0.635	0.403
(7) Life changes worth it	0.04 (0.22)	0.35	0.63	0.645	0.415
(8) I miss my baby when not together	0.10 (0.36)	0.52	0.61	0.705	0.497
(9) Feels like someone else's baby	0.04 (0.27)	0.35	0.63	0.637	0.406
(10) Look forward to seeing baby again	0.03 (0.23)	0.43	0.62	0.722	0.521
(11) Enjoy playing with	0.04 (0.23)	0.49	0.61	0.797	0.636

**Cronbach's alpha for test with all 11 item responses included was 0.64.*

TABLE 4 | Item 'endorsibility' ('measure') of the MaMB scale, along with the Rasch fit statistics.

Item	Item difficulty/'Endorsibility'	Infit (mean-squared)	Infit (standardized)	Outfit (mean-squared)	Outfit (standardized)
(1) Enjoy looking after baby	0.69	0.76	−1.85	0.53	−2.24
(2) Feel irritated with baby	−0.73	0.91	−0.67	0.65	−1.53
(3) Affectionate toward baby	0.34	0.82	−0.56	0.53	−0.75
(4) Feel baby is being difficult	0.77	0.97	−0.01	1.33	0.73
(5) Can work out baby's needs	−2.53	1.20	2.03	1.31	2.66
(6) Can't do enjoyable things because of baby	−0.11	1.03	0.29	0.95	−0.39
(7) Life changes worth it	0.53	1.07	0.34	1.31	0.75
(8) I miss my baby when not together	0.13	1.06	0.37	1.05	0.27
(9) Feels like someone else's baby	0.14	1.17	0.56	2.09	1.35
(10) Look forward to seeing baby again	0.46	0.98	0.07	0.91	0.05
(11) Enjoy playing with	0.32	0.86	−0.46	1.15	0.45

These include both 'infit' and 'outfit' statistics as both the mean squared error and standardized (z) fit.

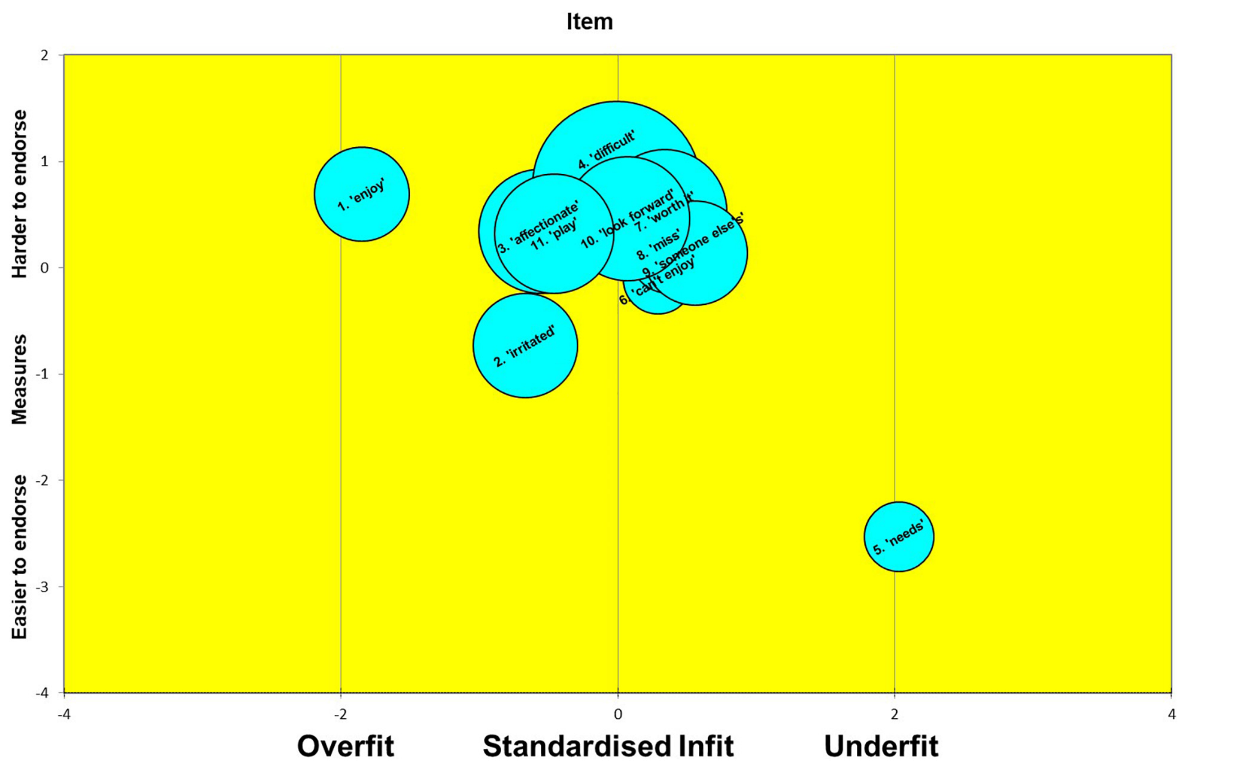

FIGURE 2 | Bubble plot of the MaMB items, according to estimated endorsability ('measure'), their standard error for this (diameter) and degree of 'infit' according to the Rasch model.

and this can be sensitive to relatively large numbers of observations (e.g., >300).

(3) One items ('I feel like I'm looking after my baby for someone else' -item 9) tends to show poor outfit. This suggests some erratic ratings, by those respondents whose estimated trait level was some distance from the item 'measure' (endorsibility').

(4) One item showed poor infit and outfit, at least on the 'z' fit statistics ('I can work out what my baby needs from me'). This suggests this item may have been relatively erratically answered. It may have been different respondents read or interpreted the item differently from each other. For example, some may have interpreted it in terms of basic needs, whilst others, more in terms of emotional needs. It may be useful to explore whether this item showed any item bias or differential item functioning in relation to demographic factors.

In terms of 'person fit'; only 16 of the 438 (3.7%) participants showed marked underfit to the Rasch model, as indicated by a standardized infit or outfit of greater than 2.0. That is, their responses were more erratic than the Rasch model would have

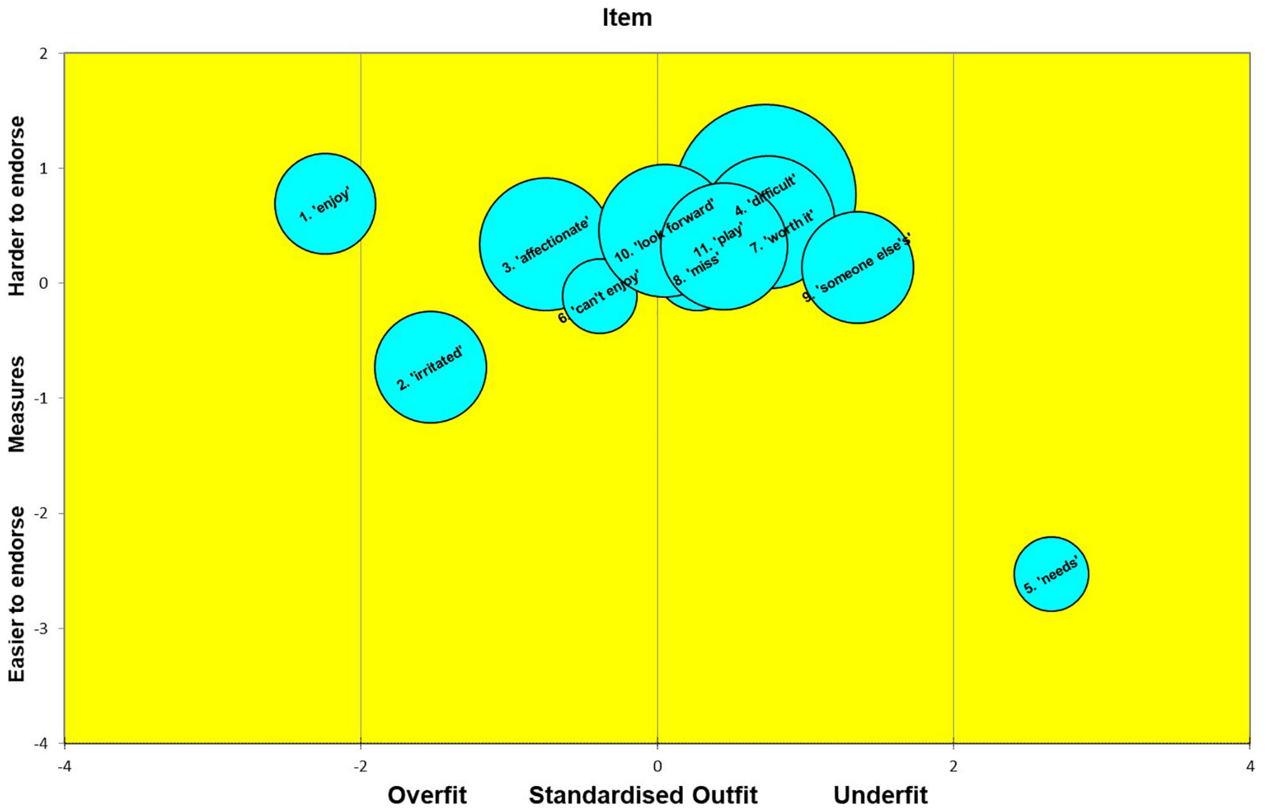

FIGURE 3 | Bubble plot of the MaMB items, according to estimated endorsability ('measure'), their standard error for this (diameter) and degree of 'outfit' according to the Rasch model.

predicted. In contrast, only one respondent showed marked overfit, as defined as a standardized infit and/or outfit of less than –2.0.

The potential for item responses to be dependent on each other was investigated by examining the matrix of correlated residuals from the Rasch model, between pairs of items. In general, the magnitude of these were very small (average 0.08). The only more substantial correlated residual ($\geq$3.0) was observed for that between item 5 ('I can work out what my baby needs from me') and item 6 ('I feel like I can't do things I enjoy because of my baby'). These two items had a correlated residual of –0.31. It is not clear why this dependency was observed, though given only one paired correlation out of 55 pairs exceeded 0.3 in magnitude this could be a chance finding.

Item Category Probabilities

It was apparent that most of the items were not operating as four-point Likert scales. That is, in many items not all four categories of response were observed in this sample of respondents. Moreover, some intermediate categories of response were rarely observed. In effect this means that even if a respondent is higher on a trait level a lower category of response may still be observed. This is sometimes referred to as 'Rasch-Andrich threshold suppression.' This effect is nicely illustrated below, by the item category probability curves for item 11. Although some respondents selected a response with a score of '2' had higher trait levels than those who scored '1' ('0' was not observed), in practice they were more likely to be seen to choose a '1' category, as so few chose the '2.' These findings suggest, at least for the kind of general population sample used in this study, the use of four Likert scale points may be too many; that is, they may not lead to more information on a test-taker and introduce some risk of extreme responses style (ERS) bias. **Supplementary Figure 1**, refers to probability of observing a respondent choosing a particular response category according to their overall trait level ('baby bonding'). Note that curves do not always correspond to the ordered responses (0→1→2→3).

Test Information

As would be expected for a test mainly composed of easily endorsed items, most of the area under the test information curve was for test takers whose traits were defined as slightly below the average. That is, those who were likely to give midrange responses to easily endorsed items. This can be seen by the fact the peak of the test information curve is just below the zero on the *x*-axis. This suggests the item calibration is not ideal to pick out mothers who may be struggling to bond with their babies (i.e., those who are likely to be observed with a lower total score on the MaMB scale). The test information curve is depicted in **Supplementary Figure 2**.

DISCUSSION

There is a paucity of high-quality tools to assess parent-infant relationships. The MaMB was co-developed to address this gap and act as a tool to measure bonding for use in research and universal health settings.

The results suggest that it is feasible for HVs to administer the MaMB with mothers in universal services. HVs successfully completed the MaMB with approximately 50% of the universal population at the 6–8-week visit in the context of highly pressured services due to the COVID-19 global pandemic. Given low rates of missing data the MaMB appears to be acceptable to parents.

The psychometric analyses suggest the MaMB tool responses, in this sample of test takers, were unidimensional. The MaMB showed relatively high levels of internal reliability consistency and the items generally fitted the Rasch model. However, the high reliability may be partly an artifact of the lack of variation in responses observed – almost all respondents gave high-scoring categories on the items. The items did not generally behave as four-point response format questions, as it was common for some response categories to go unobserved. Consequently, test information was relatively low and was much less than may be required to identify at least two separate groups of respondents, e.g., if the MaMB were to be used as a screening tool.

For the 29 parents that completed the free text it appeared a useful part of the MaMB to expand on item completion with an opportunity to voice feelings or concerns. Responses suggest parents were engaging in a meaningful discussion about bonding with their health visitor. This suggests the MaMB could be considered a potential catalyst in opening discussions about sensitive aspects of parenting such as experiencing guilt for wanting some 'alone' time, or for feeling less positive when their baby is screaming. Such open conversations suggest that the tool could fit well within a pathway for accessing specialist services, such as infant mental health services.

Strengths

The MaMB was co-developed over a series of workshops and interviews, using an iterative process with HVs, Clinical Psychologists, service staff and managers from different localities, and included parental input. It was piloted within routine HV contacts and, although the pilot was delivered during the COVID-19 pandemic with many visits taking place remotely, or with restrictions, completed MaMBs were obtained from 50% of the eligible population. The pilot study was classed as research as opposed to service design and had ethical approval as such. Previously psychometric analyses focused on exploratory and confirmatory factor analysis; however, this study also included IRT, which affords additional rigor and confidence in the results.

Limitations

Some HV teams would have conducted some core 6–8-week contacts over the telephone rather than in the family home due to COVID-19. However, we do not have data to report how many. This may have led to lower completion rate of the MaMB.

A much smaller than anticipated comparison group of non-completers was achieved. This was because HVs appeared not to complete, or partially complete, a cover sheet with demographic information if a mother did not wish to complete the MaMB. The pilot was conducted during the COVID-19 pandemic, during which time HVs were under enormous pressure to continue delivering statutory support to families despite adverse circumstances which likely contributed to the non-completion of cover sheets.

Deviations From the Registered Protocol

Due to the limited information on non-completers we were unable to conduct planned statistical analyses of the characteristics of completers compared to non-completers. The amount of data contained within the free-text responses of completed MaMBs also prevented a planned thematic analysis of these data, though it was sufficient to provide useful information in a descriptive summary.

Future Research

The findings of this study suggest that the MaMB is a promising tool to assess parent–infant relationships. Future research directions fall across three domains (1) understanding practitioner experiences, (2) expanding sample of users, and (3) refining approach to measurement.

Understanding Practitioner Experiences

Practitioners such as health visitors are a key component of using a measure of parent–infant relationships. A better understanding of their experience supporting mothers to complete the MaMB tool would help to further refine the tool. Obtaining ethical approval to ask HVs from the current study their views on completing the MaMB would be a priority for future research.

Expanding Sample of Users

This study found that most participants responded similarly to items on the MaMB. Further piloting of the tool with an expanded sample of users would help to understand the reason for this limited range of responses. For example, use with mothers experiencing mental health difficulties in the perinatal period would be particularly valuable. We might hypothesize that those within the clinical range of depression measures may respond differently when asked about their bond with their baby. This is highly likely to result in observing more variance in the items. It may also be able to show whether the tool is able to discriminate, with any precision, between at least two different groups of respondents. Note, in theory, a Rasch model is based on a sample free distribution (that is the estimates should be the same irrespective of the sample of test takers used for the calibration). However, in practice, precise estimates of item fit and difficulty may not be achieved, even with large samples, if some categories of response are rarely or never observed.

It was appropriate for this first pilot to target mothers, who are typically primary caregivers. However, we know that there is increasing variability in those who take on the primary caregiver role across society. Piloting the MaMB tool with a diverse range of caregivers would enable exploration of differences and similarities across responses for wider parent–infant relationships. It would also support use of the tool in practice, where fathers, same sex parents, or other kinship carers may be caring for a baby.

Refining Approach to Measurement

To enable the tool to have a greater degree of variation across responses, future research could test the MaMB tool with amended items (as highlighted in the results) to make them more subtle. This could be helpful in picking up difficulties and bonding and attachment in parents or caregivers. Moreover, future research could evaluate the tool as a three-point Likert scale, as opposed to the four-point scale used in the current study. This could help to increase variation across items.

CONCLUSION

Health visitors successfully administered the MaMB in universal services and the MaMB appears to be acceptable to parents. The MaMB demonstrated good internal consistency and may support HV signposting decisions for additional support, however, as the more robust analysis shows, if the MaMB was to be used as a screening tool, with a cut-off, or ranges of 'concern' then additional work is needed, which will need to include more families with risk factors such as depression in an enriched sample.

Regarding our objectives, we consider the MaMB to be feasible for use in routine practice with some amendments, and future piloting of such amendments.

ASSOCIATED PROTOCOL

Version 2, 7th December 2020 – to access the protocol for further information please visit: https://osf.io/q3hmf/.

Bywater, T., Blower, S. L., Dunn, A., Endacott, C., Smith, K., and Tiffin, P. (2020, December 7). Me and My Baby Protocol: The measurement properties and acceptability of a new mother-infant bonding tool designed for use in a universal healthcare setting in the United Kingdom. Retrieved from osf.io/6br2e.

STUDY SPONSOR

The University of York. Data sharing agreements and data management were compliant with the General Data Protection Regulation (GDPR) and University of York data management policies.

RESEARCH REFERENCE NUMBERS

- IRAS Number: 273708
- This study received ethical approval on 21st August 2020 by South Central - Berkshire B Research Ethics Committee, United Kingdom, Ref: 20/SC/0266, Integrated Research Application System (IRAS) 201, project ID: 273708.
- Funder References: HEIF - H0026802, ARC-YH: NIHR200166, The National Lottery Community Fund (previously the Big Lottery Fund) – 0094849, National Institute for Health Research (NIHR) Public Health Research (PHR) (ref 13/93/10).
- OFS Study Registration Number: osf.io/6br2e.

DATA AVAILABILITY STATEMENT

The datasets presented in this manuscript are not readily available because the data sharing agreement stipulated RDaSH anonymized data could be shared with UoY, but no agreement exists for UoY to share the data externally. Additional agreements would possibly be required to do this. Requests to access the datasets should be directed to tracey.bywater@york.ac.uk.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Berkshire REC. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

TB secured funding, wrote the initial draft, will act as guarantor and affirms that the manuscript is an honest, accurate, transparent, and full account. TB and AD conceived the study. TB, AD, CE, KS, PT, and SB designed various aspects of the study. TB and SB provided supervision of the study from the academic perspective (UoY) and KS from the practitioner perspective (RDaSH). PT conducted the statistical analysis and provided psychometric expertise. KS and MP participated in co-developing the tool and provided clinical and practitioner expertise. AD and CE were research fellows on this study and coordinated and conducted various activities. All authors have contributed and commented on subsequent drafts of this manuscript, meet authorship criteria and no others meeting the criteria have been omitted.

FUNDING

Higher Education Innovation Fund (HEIF; Ref H0026802), Research and Enterprise, University of York provided funding for some research fellow time and expenses. National Institute for Health Research; Applied Research Collaboration – Yorkshire and Humber (NIHR; ARC-YH; Ref NIHR200166) supported TB and SB and provided additional in-kind costs. National Institute for Health Research (NIHR) Public Health Research (PHR) (ref 13/93/10) funded Open Access fees and some staff time. RDaSH provided their staff time and other in-kind costs. The National Lottery Community Fund (previously the Big

Lottery Fund) funded some staff time, and other in-kind costs (Ref 10094849).

ACKNOWLEDGMENTS

This study could not be conducted without the support of RDaSH, the health visiting teams from Doncaster and North Lincolnshire, and of course the families – we extend our gratitude to these key stakeholders, and to other professionals and parents across Yorkshire and Lincolnshire who participated in the co-development of the MaMB. We would also like to thank the funders (see funding section for information), and the ARC-YH management group for their oversight.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.804885/full#supplementary-material

REFERENCES

Ainsworth, M., Blehar, M., Waters, E., and Wall, S. N. (1978). *Patterns of Attachment: A Psychological Study of the Strange Situation*. Hove, UK: Psychology Press.

Appleton, J. V., Harris, M., Oates, J., and Kelly, C. (2013). Evaluating health visitor assessments of mother-infant interactions: a mixed methods study. *Int. J. Nurs. Stud.* 50, 5–15. doi: 10.1016/j.ijnurstu.2012.08.008

Asparouhov, T., and Muthén, B. (2009). Exploratory Structural Equation Modeling. *Multidiscipl. J.* 16, 397–438. doi: 10.1080/10705510903008204

Barlow, J., McMillan, A. S., Kirkpatrick, S., Ghate, D., Barnes, J., and Smith, M. (2010). Health-Led Interventions in the Early Years to Enhance Infant and Maternal Mental Health: a Review of Reviews. *Child Adolesc. Mental Health* 15, 178–185. doi: 10.1111/j.1475-3588.2010.00570.x

Barlow, J., Schrader-McMillan, A., Axford, N., Wrigley, Z., Sonthalia, S., Wilkinson, T., et al. (2016). Review: attachment and attachment-related outcomes in preschool children – a review of recent evidence. *Child Adolesc. Mental Health* 21, 11–20. doi: 10.1111/camh.12138

Baur, T., and Lukes, D. (2009). *An Evaluation of the IRT Models Through Monte Carlo Simulation*. Madison, WI: University of Wisconsin.

Bird, P. K., Hindson, Z., Dunn, A., de Chavez, A. C., Dickerson, J., Howes, J., et al. (2021). Implementing the Maternal Postnatal Attachment Scale in universal services: qualitative interviews with health visitors. (Submitted). *medRxiv* [preprint]. doi: 10.1101/2021.11.30.21267065

Blower, S. L., Gridley, N., Dunn, A., Bywater, T., Hindson, Z., and Bryant, M. (2019). Psychometric Properties of Parent Outcome Measures Used in RCTs of Antenatal and Early Years Parent Programs: a Systematic Review. *Clin. Child Fam. Psychol. Rev.* 22, 367–387. doi: 10.1007/s10567-019-00276-2

Bowlby, J. (1982). Attachment and loss: retrospect and Prospect. *Am. J. Orthopsych.* 52, 664–678. doi: 10.1111/j.1939-0025.1982.tb01456.x

Branjerdporn, G., Meredith, P., Strong, J., and Garcia, J. (2017). Associations Between Maternal-Foetal Attachment and Infant Developmental Outcomes: a Systematic Review. *Maternal Child Health J.* 21, 540–553. doi: 10.1007/s10995-016-2138-2

Brockington, I. F., Fraser, C., and Wilson, D. (2006). The Postpartum Bonding Questionnaire: a validation. *Arch. Womens Ment. Health* 9, 233–242. doi: 10.1007/s00737-006-0132-1

Bywater, T. J., Berry, V., Blower, S. L., Cohen, J., Gridley, N., Kiernan, K., et al. (2018). Enhancing social-emotional health and wellbeing in the early years (E-SEE). A study protocol of a community-based randomised controlled trial with process and economic evaluations of the Incredible Years infant and toddler parenting programmes, delivered in a Proportionate Universal Model. *BMJ Open* 8:e026906. doi: 10.1136/bmjopen-2018-026906

Cassidy, J., Jones, J. D., and Shaver, P. R. (2013). Contributions of attachment theory and research: a framework for future research, translation, and policy. *Devel. Psychopathol.* 25, 1415–1434. doi: 10.1017/s0954579413000692

Condon, J. T. (1993). The assessment of antenatal emotional attachment: development of a questionnaire instrument. *Br. J. Med. Psychol.* 66, 167–183. doi: 10.1111/j.2044-8341.1993.tb01739.x

Condon, J. T., and Corkindale, C. J. (1998). The assessment of parent-to-infant attachment: development of a self-report questionnaire instrument. *J. Reproduct. Infant Psychol.* 16, 57–76. doi: 10.1080/02646839808404558

Cooper, C. (2019). *Psychological Testing: Theory and Practice*. London: Routledge.

de Vet, H. W. W., Terwee, C. B., Mokkink, L. B., and Knol, D. L. (2015). *Measurement in Medicine*. Cambridge: Cambridge University Press.

Dickerson, J., Bird, P. K., McEachan, R. R., Pickett, K. E., Waiblinger, D., Uphoff, E., et al. (2016). Born in Bradford's Better Start: an experimental birth cohort study to evaluate the impact of early life interventions. *BMC Public Health* 16:711. doi: 10.1186/s12889-016-3318-0

Dunn, A., Bird, P., Bywater, T., Dickerson, J., and Howes, J. (2021). *Implementing the Maternal Postnatal Attachment Scale (MPAS) in universal services: Qualitative interviews with health visitors. medRxiv* [preprint].

Elmer, J. R. S., O'Shaughnessy, R., Bramwell, R., and Dickson, J. M. (2019). Exploring health visiting professionals' evaluations of early parent-infant interactions. *J. Reproduct. Infant Psychol.* 37, 554–565. doi: 10.1080/02646838.2019.1637831

Facompre, C. R., Bernard, K., and Waters, T. E. (2018). Effectiveness of interventions in preventing disorganized attachment: a meta-analysis. *Devel. Psychopathol.* 30, 1–11. doi: 10.1017/s0954579417000426

Fearon, R. P., Bakermans-Kranenburg, M. J., van Ijzendoorn, M. H., Lapsley, A. M., and Roisman, G. I. (2010). The significance of insecure attachment and disorganization in the development of children's externalizing behavior: a meta-analytic study. *Child Dev.* 81, 435–456. doi: 10.1111/j.1467-8624.2009.01405.x

Feldman, R., Weller, A., Leckman, J. F., Kuint, J., and Eidelman, A. I. (1999). The nature of the mother's tie to her infant: maternal bonding under conditions of proximity, separation, and potential loss. *J. Child Psychol. Psychiatry. Allied Discipl.* 40, 929–939. doi: 10.1111/1469-7610.00510

Goldman, S. H., and Raju, N. S. (1986). Recovery of One- and Two-Parameter Logistic Item Parameters: an Empirical Study. *Educat. Psychol. Measur.* 46, 11–21. doi: 10.1177/0013164486461002

Gridley, N., Blower, S., Dunn, A., Bywater, T., and Bryant, M. (2019). Psychometric Properties of Child (0-5 Years) Outcome Measures as used in Randomized Controlled Trials of Parent Programs: a Systematic Review. *Clin. Child Fam. Psychol. Rev.* 22, 388–405. doi: 10.1007/s10567-019-00277-1

Horn, J. L. (1965). A Rationale And Test For The Number Of Factors In Factor Analysis. *Psychometrika* 30, 179–185. doi: 10.1007/bf02289447

Kane, K. (2017). *Assessing the Parent-Infant Relationship Through Measures of Maternal-Infant Bonding. A Concept Analysis and Systematic Review*. Ph D thesis, England, UK: University of York.

Kristensen, I. H., Trillingsgaard, T., Simonsen, M., and Kronborg, H. (2017). Are health visitors' observations of early parent-infant interactions reliable? A cross-sectional design. *Infant Ment. Health J.* 38, 276–288. doi: 10.1002/imhj.21627

Le Bas, G. A., Youssef, G. J., Macdonald, J. A., Rossen, L., Teague, S. J., Kothe, E. J., et al. (2019). The role of antenatal and postnatal maternal bonding in infant development: a systematic review and meta-analysis. *Soc. Devel.* [Epub online ahead of print] doi: 10.1111/sode.12392

Linacre, J. (1994). *Sample Size and Item Calibration (or Person Measure) Stability*. Available online at: https://www.rasch.org/rmt/rmt74m.htm [accessed on Oct27, 2021]

Linacre, J. M. (2017). *Winsteps Rasch Measurement Computer Programme (Version 4.0.0)*. Available online at: https://www.winsteps.com/

Lorenzo-Seva, U., and Ferrando, P. J. (2006). FACTOR: a computer program to fit the exploratory factor analysis model. *Behav. Res. Methods* 38, 88–91. doi: 10.3758/bf03192753

Lorenzo-Seva, U., and Van Ginkel, J. R. (2016). Multiple imputation of missing values in exploratory factor analysis of multidimensional scales: estimating latent trait scores. *Anal. Psicol. Anna. Psychol.* 32, 596–608.

Marsh, H. W. (1996). Positive and negative global self-esteem: a substantively meaningful distinction or artifactors? *J. Pers. Soc. Psychol.* 70, 810–819. doi: 10.1037//0022-3514.70.4.810

Müller, M. E. (1994). A questionnaire to measure mother-to-infant attachment. *J. Nurs. Meas.* 2, 129–141.

NICE (2012). *Social and Emotional Wellbeing: Early Years*. France: NICE.

NICE. (2015). *Children's Attachment: Attachment in Children and Young People Who Are Adopted from Care, in Care or at High Risk of Going Into Care*. France: NICE.

Public Health England [PHE] (2020). *Early Years High Impact Area Two: Supporting Good Parental Mental Health*. London: Public Health England.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: The University of Chicago Press.

Raykov, T. (1997). Scale Reliability, Cronbach's Coefficient Alpha, and Violations of Essential Tau-Equivalence with Fixed Congeneric Components. *Multiv. Behav. Res.* 32, 329–353. doi: 10.1207/s15327906mbr3204_2

Redshaw, M., and Martin, C. (2013). Babies, 'bonding' and ideas about parental 'attachment'. *J. Reproduct. Infant Psychol.* 31, 219–221.

Taylor, A., Atkins, R., Kumar, R., Adams, D., and Glover, V. (2005). A new Mother-to-Infant Bonding Scale: links with early maternal mood. *Arch. Womens Ment. Health* 8, 45–51. doi: 10.1007/s00737-005-0074-z

Wilson, P., Thompson, L., Puckering, C., McConnachie, A., Holden, C., Cassidy, C., et al. (2010). Parent-child relationships: are health visitors' judgements reliable? *Commun. Pract.* 85, 22–25.

Wittkowski, A., Vatter, S., Muhinyi, A., Garrett, C., and Henderson, M. (2020). Measuring bonding or attachment in the parent-infant-relationship: a systematic review of parent-report assessment measures, their psychometric properties and clinical utility. *Clin. Psychol. Rev.* 82:101906. doi: 10.1016/j.cpr.2020.101906

Wright, B., Barry, M., Hughes, E., Trépel, D., Ali, S., Allgar, V., et al. (2015). Clinical effectiveness and cost-effectiveness of parenting interventions for children with severe attachment problems: a systematic review and meta-analysis. *Health Technol. Assess.* 19:52. doi: 10.3310/hta19520

Wright, B., and Masters, G. (1982). *Rating Scale Analysis*. Chicago, IL: MESA Press.

Author Disclaimer: The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. This study also received funding from the National Lottery Community Fund (previously the Big Lottery Fund) as part of the 'A Better Start' program. The National Lottery Community Fund have not had any involvement in the design or writing of the manuscript.

Conflict of Interest: TB and SB are supported by the NIHR Yorkshire and Humber Applied Research Collaboration (ARC–YH). KS is an employee of RDaSH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 16 February 2022
doi: 10.3389/fpsyg.2022.800285

Cross-Cultural Adaptation and Psychometric Evaluation of the Perceived Ability to Cope With Trauma Scale in Portuguese Patients With Breast Cancer

*Raquel Lemos[1,2], Beatriz Costa[1,3], Diana Frasquilho[4], Sílvia Almeida[1,5], Berta Sousa[1,4,6] and Albino J. Oliveira-Maia[1,3]**

[1] Champalimaud Research and Clinical Centre, Champalimaud Foundation, Lisbon, Portugal, [2] ISPA — Instituto Universitário de Ciências Psicológicas, Sociais e da Vida, Lisbon, Portugal, [3] NOVA Medical School, NMS, Universidade Nova de Lisboa, Lisbon, Portugal, [4] Breast Unit, Champalimaud Clinical Centre, Champalimaud Foundation, Lisbon, Portugal, [5] Graduate Programme in Clinical Psychology, Faculdade de Psicologia da Universidade de Lisboa, Lisbon, Portugal, [6] Ph.D Programme in Health Data Science, Faculdade de Medicina da Universidade do Porto, Porto, Portugal

OPEN ACCESS

Edited by:
*Paola Gremigni,
University of Bologna, Italy*

Reviewed by:
*Ana Galhardo,
Instituto Superior Miguel Torga, Portugal
Nancy Van Loey,
Association of Dutch Burn Centres, Netherlands*

***Correspondence:**
*Albino J. Oliveira-Maia
albino.maia@neuro.fchampalimaud.org*

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *22 October 2021*
Accepted: *13 January 2022*
Published: *16 February 2022*

Citation:
Lemos R, Costa B, Frasquilho D, Almeida S, Sousa B and Oliveira-Maia AJ (2022) Cross-Cultural Adaptation and Psychometric Evaluation of the Perceived Ability to Cope With Trauma Scale in Portuguese Patients With Breast Cancer. Front. Psychol. 13:800285. doi: 10.3389/fpsyg.2022.800285

Background: The impact of a cancer diagnosis may be traumatic, depending on the psychological resources used by patients. Appropriate coping strategies are related to better adaptation to the disease, with coping flexibility, corresponding to the ability to replace ineffective coping strategies, demonstrated to be highly related with self-efficacy to handle trauma. The Perceived Ability to Cope with Trauma (PACT) scale is a self-rated questionnaire that assesses the perceived ability to cope with potentially traumatic events, providing a measure of coping flexibility. The current study aimed at examining the psychometric properties of the PACT Scale in Portuguese patients with breast cancer.

Methods: The study included 172 patients recently diagnosed with early breast cancer. Participants completed a Portuguese version of the PACT scale, and instruments of self-efficacy for coping with cancer (Cancer Behavior Inventory-Brief Version—CBI-B), of quality of life (European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core-30—QLQ-C30), and of psychological distress (Hospital Anxiety and Depression Scale—HADS) that were used as convergent and divergent measures, thus assessing construct validity. A confirmatory factor analysis (CFA) was performed to test the factor structure of the Portuguese version of PACT scale and reliabilities were examined.

Results: Results from the CFA confirmed the two-factor structure, consistent with the original Forward and Trauma focus subscales. The two subscales demonstrated high internal consistencies. Convergent and divergent validities were confirmed: the PACT scale was related to high self-efficacy to cope with cancer (CBI-B), to high perceived quality of life (QLQ-C30), and to low psychological distress (HADS).

Discussion: Overall, the current results support and replicate the psychometric properties of the PACT scale. The scale was found to be a valid and reliable self-reported measure to assess Portuguese breast cancer patients regarding beliefs about their capabilities in managing the potentially traumatic sequelae of cancer. The PACT is a simple and brief measure of coping flexibility to trauma, with potential relevance for application in clinical and research settings.

Keywords: cross-cultural adaptation, validity, psychometrics, cancer, trauma, coping flexibility

INTRODUCTION

The burden of cancer incidence in 2020 is estimated to have risen to 19.3 million new cases worldwide (Sung et al., 2021). Female breast cancer was the leading form of cancer globally in 2020, with an estimated number of 2.3 million new cases worldwide, representing 11.7% of all cancer cases (Sung et al., 2021). Among women, breast cancer accounts for 1 in 4 cancer cases, ranking first for incidence in the majority of countries (Sung et al., 2021). According to the World Health Organization Global Cancer Observatory, in Europe breast cancer was responsible for 12.1% of all new cancer cases in 2020, representing 25.8% of all female cancers.[1] In Portugal, for example, the number of female breast cancer cases in 2020 was 7,041, representing 26.4% of cancer diagnoses in women (see text footnote 1).

Being diagnosed with cancer and experiencing cancer treatment is highly stressful, with the potential to become a traumatic experience, threatening physical and psychological wellbeing. Emotional reaction to this experience includes acute responses of fear, sadness, and anger, but also long-term adjustment difficulties characterized by anxiety and depression (Foster et al., 2009). Several stressors are associated with cancer diagnosis and the treatment trajectory. Uncertainty about prognosis, management of clinical information and decision-making regarding treatments can make the early phases of the cancer path particularly overwhelming (Hack et al., 2010). In the specific case of female breast cancer, the diagnosis may additionally challenge identity, self-esteem, body image and relationships (Campbell-Enns and Woodgate, 2015). Women who experienced distress due to breast cancer are at higher risk of feeling an impact on long-term quality of life, with estimations of 20–30% of survivors reporting psychological difficulties that persist for years after the diagnosis (Foster et al., 2009). Nevertheless, while a considerable proportion of people may experience cancer diagnosis and treatments as traumatic, this is not true for everyone (for meta-analytic reviews see Abbey et al., 2015; Swartzman et al., 2017).

These findings led to research on the prevalence, predictor factors and correlates of cancer-related post-traumatic stress disorder (PTSD) symptoms (Foster et al., 2009). Facing a cancer diagnosis differs from typical acute traumatic events: the stressor comes from an internal, rather than an external, locus, and individuals deal constantly with the presence of an ongoing threat, as opposed to experiencing a single past-incident traumatic event (Pat-Horenczyk et al., 2016). Studies exploring cancer as a traumatic stressor have used the DSM-IV-TR (American Psychiatric Association, 2000) PTSD diagnostic criteria to understand if patients experienced cancer diagnosis and treatments as a threat to their life or physical integrity (criterion A1) and if they reacted with fear, helplessness or horror (criterion A2). Across studies, 50–60% of patients endorsed the two criteria, with the first criterion endorsed more commonly than the second (Cordova et al., 2001, 2007, 2017). Among patients with cancer, Matthews et al. (2017) found that PTSD symptoms were more frequently reported by women than men (27% vs. 10%). The predictors of PTSD among women included perceived intensity of cancer treatment, difficulties with health care professionals, and using cognitive avoidant coping styles. For men the only predictor was behavioral avoidance (Matthews et al., 2017). Mehnert and Koch (2007) reported that cancer was a traumatic stressor for 54% patients with breast cancer, with patients scoring high in avoidant symptoms just after receiving the diagnosis presenting difficulties in adjustment up to 2 years later (Arnaboldi et al., 2017).

Since the impact of a cancer diagnosis varies according to the psychological resources that patients use, understanding how patients cope and adjust to a cancer diagnosis is essential to plan care. Macía et al. (2020) showed that, in people with cancer, the use of appropriate coping strategies and the presence of higher levels of resilience were related to better quality of life and better adaptation to the disease. The ability to engage in adaptive coping behaviors predicts optimal adjustment in the presence of highly aversive or potentially traumatic life events (Bonanno, 2004, 2005), such as after receiving a cancer diagnosis. Moreover, while active or instrumental coping strategies, such as positive thinking or dealing actively with problems, are associated with a positive adaptation to stress, passive coping strategies (i. e., avoidance) are usually considered maladaptive (for a review see Linley and Joseph, 2004; Cheng et al., 2014). Understanding the relationship of resilience and coping with quality of life represents valuable information for psychologists working in the oncological setting. At a practical level, it may implicate working with patients in modifying the type of coping, as well as increasing the level of resilience, toward achieving better adjustment to cancer.

In the specific case of breast cancer, self-efficacy to cope with cancer tends to improve over time after diagnosis (Kochaki Nejad et al., 2015), and has been associated to many well-being outcomes through a combination of cognitive, emotional, and behavioral variables (Karademas et al., 2021). Furthermore, coping self-efficacy has been shown to mediate the relationship between illness perception and fear of progression (Shim et al., 2018) as well as between perceived social constraints and symptoms among long-term survivors (Adams et al., 2017). Kant et al. (2018) observed that coping self-efficacy following breast cancer diagnosis predicts less psychological distress over time and

[1] https://gco.iarc.fr

a recent review confirmed that self-efficacy can predict quality of life and psychological distress in patients with breast cancer, thus highlighting that, at the time of diagnosis, it is important to identify women at risk for psychological distress (Brandão et al., 2017). Among coping strategies used by these patients, there have been reports of the importance of accepting the diagnosis and engaging in physical activities providing social and emotional support (Lashbrook et al., 2018).

Coping research has recently suggested the concept of "coping flexibility" to overcome the lack of diversity and fluidity of coping, as considered in more classical research (Kato, 2012). This concept is based on the transactional theory presuming that coping may change over time according to the demands of a specific stressful situation (Lazarus and Folkman, 1984). As such, coping flexibility corresponds to the ability to discontinue an ineffective coping approach and produce and implement an alternative one (Kato, 2012). Accordingly, in the context of trauma, resilience would be fostered by the ability to flexibly engage in different coping strategies as needed, and not by a single type of coping (Bonanno, 2004, 2005; Bonanno et al., 2011).

According to a review on coping flexibility (Cheng et al., 2014), the most widely used instruments for measuring this construct are the self-rated Flexible Goal Adjustment Scale (FGAS; Brandtstädter and Renner, 1990) and the Coping Flexibility Questionnaire (CFQ; Cheng, 2001). The FGAS is a 15-item questionnaire, rated in a 5-point Likert scale, to assess the ability to modify coping goals according to changing environments. It assesses coping according to two perspectives: *assimilation*, i.e., seeking to change one's development conditions according to personal preferences (*Tenacious Goal Pursuit* subscale); and *accommodative flexibility*, i.e., indicating the adjustment of individual preferences to situational limits (*Flexible Goal Adjustment* subscale). Yet, studies using this scale have noted that it fails to adequately distinguish between its two subscales (Henselmans et al., 2011). The CFQ is an open-ended, situation-based measure. It assesses, through a 6-point Likert scale, coping responses to a series of stressful life events. Flexible coping is obtained as an individual coping profile, indicating the frequency of using different strategies according to each stressful situation. Limitations of this scale have been shown in cross-cultural studies, with differences in the use of coping strategies according to culture (Basińska et al., 2021). Other scales have also been proposed to assess flexible coping. The Coping Flexibility Scale (CFS; Kato, 2012) is based on the dual-process model of coping flexibility previously proposed in the FGAS. The flexible goal adjustment process was refined by suggesting an additional process that precedes it. According to Kato (2012) one should be able to recognize that a strategy no longer works before implementing an alternative (i.e., adaptive coping process). The CFS assesses two flexible coping processes across the *Evaluation coping* and *Adaptive coping* subscales. Similar to the limitations of CFQ, the CFS has shown to be susceptible to cultural influences (Basińska, 2015). The Coping Flexibility Questionnaire (COFLEX; Vriezekolk et al., 2012) is a 13-item instrument that includes two dimensions of coping flexibility: *versatility*, the capability of using the different available coping resources according to the circumstances, and *reflective coping*, the capability of generating and considering coping options, and estimating the suitability of a coping strategy in a given situation. While there was preliminary evidence of the validity of the versatility dimension, for reflective coping it could not be firmly established. Moreover, Basińska et al. (2021) raised another weakness of the scale, based on the fact that its development and the selection of items was guided by theory and selected by researchers, instead of being produced by patients.

In the specific context of potentially traumatic events, Bonanno et al. (2011) examined the notion of flexible coping according to concepts of perceived ability. They hypothesized that effectively coping with trauma involves the flexible use of two coping processes: *forward focus*, the perceived ability to move beyond the trauma, and *trauma focus*, the perceived ability to process the trauma. These two coping strategies are assessed by the Perceived Ability to Cope with Trauma scale (PACT; Bonanno et al., 2011). This is a self-rated questionnaire, explicitly designed to assess the perceived ability to cope with potentially traumatic events, providing a measure of coping flexibility. The PACT displayed adequate reliability and validity across Israeli and American samples (Bonanno et al., 2011), demonstrating cross-cultural adequacy. It has been mainly used in studies of coping in highly trauma-exposed samples (Bonanno et al., 2011; Park et al., 2015; Bartholomew et al., 2017; Pinciotti et al., 2017; Sullivan and Wade, 2019), in potentially traumatic life events such as adjustment to college (Bonanno et al., 2011; Galatzer-Levy et al., 2012; Saita et al., 2017), grief severity in older widows and widowers (Knowles and O'Connor, 2015), experience with the COVID-19 pandemic (Zhou et al., 2020; Brivio et al., 2021), and in breast cancer patients (Hamama-Raz et al., 2012; Pat-Horenczyk et al., 2016; Brivio et al., 2021). The present study aimed to adapt the European Portuguese version of the PACT scale and validate it for patients with breast cancer.

MATERIALS AND METHODS

Participants

This study was conducted at the Champalimaud Clinical Centre under the multicenter clinical study—BOUNCE (Predicting Effective Adaptation to Breast Cancer to Help Women to BOUNCE Back).[2] This study included the completion of questionnaires assessing quality of life, mood, and personal characteristics. Patients with a recent diagnosis of stage I-III histologically confirmed breast cancer, and eligible for systemic treatment, were recruited at their first clinical visit to the oncologist and invited to participate in the study before starting any systemic treatment. Eligibility criteria included: women 18–70 years of age at the time of diagnosis, histologically confirmed invasive breast cancer, tumor stages I—III, local treatment with surgery with or without adjuvant radiation therapy, any type of systemic treatment. Exclusion criteria were: distant metastasis; history of another malignancy or contralateral invasive breast cancer within the last 5 years, with the exception of cured basal cell carcinoma of skin or carcinoma *in situ* of uterine

[2] https://www.bounce-project.eu/

cervix; history of early onset (i.e., < 40 years of age) mental disorder (i.e., schizophrenia, psychosis, bipolar disorder, major depression) or severe neurologic disorder (i.e., neurodegenerative disorder); other serious concomitant diseases such as clinically significant (i.e., active) cardiac disease (e.g., congestive heart failure, symptomatic coronary artery disease or uncontrolled cardiac arrhythmia) or myocardial infarction within the last 12 months; major surgery for a severe disease or trauma which could affect patient's psychosocial wellbeing (e.g., major heart or abdominal surgery) within 4 weeks prior to study entry, or lack of complete recovery from surgery.

Measures

Sociodemographic and Lifestyle Questionnaire and Medical Data

This questionnaire was developed specifically for this study to assess sociodemographic and lifestyle variables, such as age, educational level, marital status, and employment status, as reported by the participants.

Perceived Ability to Cope With Trauma

The PACT Scale (Bonanno et al., 2011) is a 20-item self-report measure of beliefs about the capability to manage traumatic sequelae. Answers are given in a Likert-type scale that ranges from 1 ("Not at all able") to 7 ("Extremely able"). The original version indicated the presence of two subscales: Forward focus and Trauma focus. Forward focus (12 items) was identified by the authors as the perceived ability to move beyond the trauma, i.e., assessing coping abilities related to keeping plans and goals, attending to the needs of others, thinking optimistically, remaining calm, reducing painful emotions, and being able to laugh. The Trauma Focus subscale (8 items) was proposed to measure the perceived ability to process the trauma, through a full experience of the emotional and cognitive significance of a stressful and potentially traumatic event. An algorithm index of flexibility is calculated to estimate the ability to engage in both types of coping (Bonanno et al., 2011). The coping flexibility score is computed by (1) adding the average scores of the trauma focus and the forward focus subscales to create a total coping score, (2) creating a polarity score by taking the absolute values of the difference between the standardized trauma focus and forward focus subscale scores, and (3) subtracting the polarity score from the total coping score to generate a coping flexibility score. Greater scores mean greater coping flexibility strategies. In the current study, analyses will cover the two PACT scales (Trauma Focus and Forward Focus), as well as the total coping and the flexibility scores. In the original version, the Cronbach alpha was 0.91 for the Forward Focus and 0.79 for the Trauma Focus (Bonanno et al., 2011).

Cancer Behavior Inventory-Brief Version

The Cancer Behavior Inventory-Brief Version [CBI-B; (Heitzmann et al., 2011); Portuguese Version by Pereira et al. (2021)] is a brief version derived from the Cancer Behavior Inventory-Long (33 items). The instrument consists of 12 items and is a measurement of self-efficacy for behaviors related to coping with cancer. Following each item is a Likert-type scale that ranges from 1 ("not at all confident") to 9 ("totally confident"), reflecting the degree of confidence that cancer related coping behaviors will be performed. The total score of the scale is obtained by summing the scores of all items, where higher scores refer to higher self-efficacy in coping with cancer. In the Portuguese validation study, the Cronbach alpha was 0.88 (Pereira et al., 2021) while in the present study it was 0.86, confirming high internal consistency.

European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30

The European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core-30 [EORTC QLQ-C30; (Aaronson et al., 1993); Portuguese Version by Pais-Ribeiro et al. (2008)] is 30-item questionnaire to assess health-related quality of life (QoL) in patients with cancer, from the moment of diagnosis to long-term survivorship. The questionnaire combines five functional subscales (physical, role, emotional, cognitive, and social), three symptom subscales (fatigue, nausea and vomiting, and pain), a global health/QoL subscale, and a few single items assessing other symptoms frequently reported by cancer patients (dyspnea, insomnia, appetite loss, constipation, and diarrhea) as well as the perceived financial impact of the disease. All items are scored in 4-point Likert-type scales ranging from 1—"not at all" to 4—"very much," except two items of the global health/QoL subscale, that use a modified 7-point linear analog scale (from 1—"poor" to 7—"excellent"). Each multi-item scale includes a different set of items, i.e., no item appears in more than one scale. Higher scores received from the global health/QoL subscale indicate higher quality of life, whereas higher scores obtained from the functional or the symptom scales/items indicate lower quality of life. For this study, namely for assessing convergent validity, we only used the global quality of life score. In the Portuguese validation study, the Cronbach's alpha for the global quality of life was 0.88 (Pais-Ribeiro et al., 2008), and 0.87 in the present study.

Hospital Anxiety and Depression Scale

The Hospital Anxiety and Depression Scale [HADS; (Zigmond and Snaith, 1983); Portuguese Version by Pais-Ribeiro et al. (2007)] is a 14-item self-report measure designed to assess severity of depression and anxiety symptoms, in two separate subscales. Items are answered in a 4-point Likert-type scale response category (ranging between 0 and 3). Higher scores in the total scale indicate greater psychological distress. In the original Portuguese validation study (Pais-Ribeiro et al., 2007), including patients with cancer, Cronbach's alpha was 0.76 for the anxiety subscale and 0.81 for the depression subscale. In this study, only the total scale was calculated with a Cronbach's alpha of 0.89.

Procedures

Permissions to translate the PACT scale were obtained from the original authors by the BOUNCE study project manager. We then followed the International Test Commission Guidelines for Translating and Adapting Tests (Gregoire, 2018). Briefly, forward translation of the original scales from English to European Portuguese was performed separately by two bilingual experts in Psychology of Portuguese dominant language, resulting in two forward translated versions of the scale. A translation panel composed of mental health and oncology specialists who had not been involved in any of the translations then conducted a reconciliation of the two forward translations. The reconciled translation was then backward translated into English by two bilingual translators, of English dominant language, that worked independently and were not involved in the original translations. This was followed by comparison of the backward translated versions by the translation panel, thus creating a consensus backward translation version, that was compared against the source language by the initial translation team, to confirm similarity between the two versions and address any potential major differences in the consensus backward translation by adjustments of the consensus forward translation. The resulting harmonized version of the consensus forward translation was tested among representatives of the target population and language group (6 Portuguese patients with breast cancer), in a cognitive debriefing session to determine if the respondents understood the questions being asked and if there were words or phrases that were not familiar. No significant difficulties were reported in the debriefing session. The input from these patients considered only the replacement of some words for synonyms with a higher frequency in European Portuguese, so that it could be easily understood. For example, "*evento*" (Portuguese word for event) was replaced by "*acontecimento*" (Portuguese word for happening), as event may indicate a special moment like a party or wedding. This happened both in the test instructions and in some items. After the input from these patients, the translation was reviewed, and proofreading was conducted to ensure that minor errors were corrected, resulting in the final Portuguese translation of the scale (for an overview see **Figure 1**).

Study procedures and protocol were reviewed and approved by the Ethics Committee of our institution. Informed consent was obtained from all participants, and the study was conducted in accordance with the Declaration of Helsinki. Data was collected, stored, and processed in accordance with ethical principles and applicable international, EU and National legislation, in particular the General Data Protection Regulation (EU GDPR).

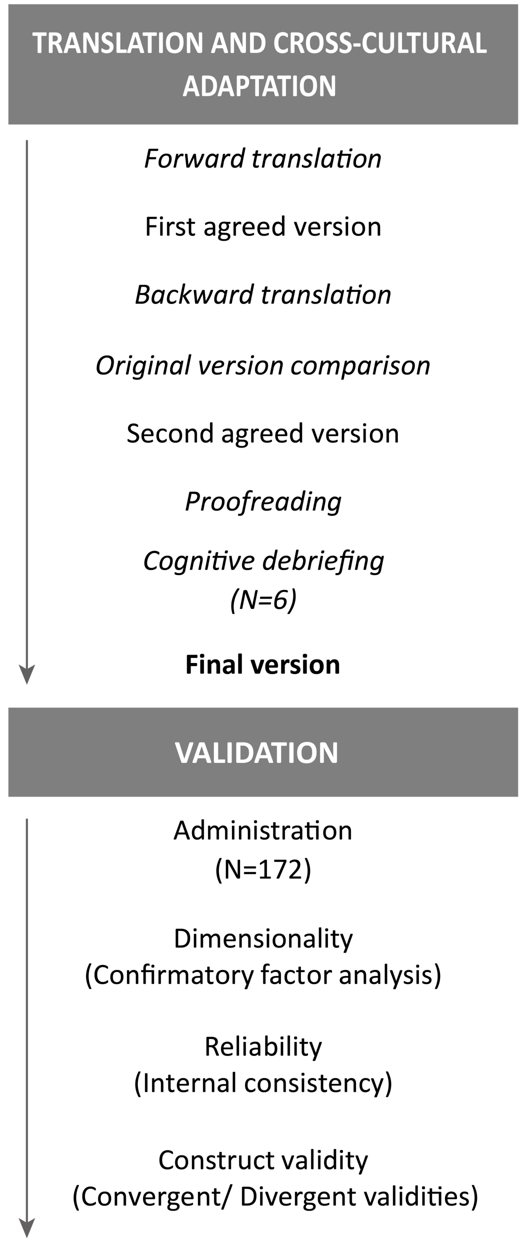

FIGURE 1 | Flowchart of the cross-cultural adaptation and validation of the Perceived Ability to Cope with Trauma (PACT).

Data Analysis

Descriptive statistics for sociodemographic and psychometric data included means and standard deviations (SD), minimum and maximum absolute values, percentages, skewness and kurtosis, that obtained using the Statistical Package for the Social Sciences (SPSS, Version 27.0; IBM SPSS, Inc., Chicago, IL). To assess dimensionality, confirmatory factor analysis (CFA) was calculated using JASP version 0.14.1,[3] which is built on the R-package lavaan.[4] A model was specified according to the original two-factor structure (Bonanno et al., 2011). Diagonally Weighted Least Squares estimation method was employed because of the ordinal scale structure and because of the relatively small sample size, according to current recommendations (Jöreskog and Sörbom, 1989; Li, 2016; Gana and Broc, 2018). Model fit indices included: (a) non-significant χ^2; (b) the comparative fit index (CFI) and the Tucker–Lewis index (TLI), with values $\geq$0.90 and $\geq$0.95 indicating good and very good model fit, respectively; and (c) the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR) indexes, with values $\leq$0.08 and $\leq$0.05 indicating acceptable and very good model fit, respectively (Hair, 2011; Gana and Broc, 2018). Item local adjustment was analyzed through the inspection of factor loadings (λ), that represent the strength of the relationship among the latent

[3]https://jasp-stats.org

[4]http://lavaan.ugent.be

variable and the observed variable. Significant ($p \leq 0.05$) factor loadings with $\lambda \geq 0.40$ are considered as a good indicator of the quality of the items (Gana and Broc, 2018). Reliability was examined using the McDonald's omega and the Cronbach's alpha, with coefficients $\geq$0.70 suggesting good factor reliability (Hair, 2011), and using the corrected item-total correlation, with values above 0.30 suggesting good inter-item correlation (Nunnally and Bernstein, 1994). To assess construct validity, we used Pearson's correlation coefficients between the PACT scores and scores on measures of self-efficacy (CBI-B) and QoL (QLQ-C30) for convergent validity, and psychological distress (HADS) for divergent validity. Independent samples t-tests were used to compare means between groups. Results with alpha-level (p) < 0.05 were considered statistically significant.

RESULTS

Descriptive Statistics

Our sample included 172 women with early or locally advanced, non-metastatic breast cancer, treated with either chemotherapy ($n = 99$) or endocrine therapy ($n = 73$), for whom sociodemographic data and clinical characteristics are presented in **Table 1**. The overall mean age ($\pm$SD) was 50.7 ($\pm$9.1), but the endocrine therapy group (53.4 $\pm$ 9.3) was significantly older than the chemotherapy group [48.7 $\pm$ 8.4; $t_{(170)} = -3.47$, $p = 0.001$], as expected. Most patients had completed higher education (47.1% with a bachelor's degree and 26.7% with a graduate degree); 74.4% were married; and 83.1% were employed.

TABLE 1 | Sociodemographic and Clinical characteristics of the sample.

Demographic and clinical characteristics (n = 172)	n	%
Age, mean (SD)	50.66 (9.08) [Min. (22); Max. (70)]	
Age group		
≤40 y	23	13.4
41–50 y	73	42.4
51–60 y	47	27.3
>60 y	29	16.9
Highest level of education		
Primary	5	2.9
Lower secondary	9	5.2
Higher secondary	31	18.0
Post-secondary non-graduate	81	47.1
Graduate degree	46	26.7
Marital status		
Single/Engaged	20	11.6
Married	128	74.4
Divorced/widowed	24	14.0
Employment status		
Employed	143	83.1
Unemployed/housewife	11	6.4
Retired	18	10.5
Treatment		
Chemotherapy (CT)	99	57.6
Endocrine Therapy (ET)	73	42.4

Descriptive statistics (means, standard deviations, kurtosis, skewness) of individual PACT items are shown in **Table 2**. The percentage of endorsement for each item is also provided, showing an overall tendency for higher value ratings (7—"extremely able").

Dimensionality

A CFA of the PACT two-factor model suggested for the original version was conducted. The general model indicated a good model fit through adequate goodness-of-fit indices: $\chi^2(169) = 166.3$, $p = 0.54$; CFI = 1.00; TLI = 1.00; and RMSEA = 0.00, 90% CI: 0.00–0.03; SRMR = 0.08. Overall, all items presented good local adjustment, with loadings ranging from $\lambda = 0.43$ (item 11) to $\lambda = 0.82$ (item 14), except for item 10 that presented a loading of 0.35 (**Figure 2**). We decided to retain the item as our model had showed an overall very good fit of the model, that would not improve with elimination of item 10.

Reliability

Internal consistency was estimated using using the McDonald's omega and the Cronbach's alpha. Factor 1 ("Forward focus" scale) showed an excellent reliability ($\omega = 0.91$, 90% CI: 0.89–0.92; $\alpha = 0.90$, 90% CI: 0.88–0.92). The values remained stable with removal of any item (ω: 0.89–0.91; α: 0.89–0.90), and corrected item-total correlations ranged between 0.51 and 0.75. Lower, but good, values were found for Factor 2 ("Trauma focus" scale): $\omega = 0.82$, 90% CI: 0.79–0.86; $\alpha = 0.83$, 90% CI: 0.79–0.86. Again, the values remained stable with removal of any item (0.79–0.82), and corrected item-total correlations ranged between 0.48 and 0.68.

Construct Validity

Table 3 shows correlations between the four PACT scores and other self-report measures selected to test convergent and divergent validity. When compared with the CBI-B scale, the PACT scores had adequate convergent validity (Factor 1: $r = 0.35$, $p < 0.001$; Factor 2: $r = 0.32$, $p < 0.001$; total coping: $r = 0.41$, $p < 0.001$; flexibility: $r = 0.26$, $p = 0.01$). The QoL scale (QLQ-C30) was also adequately correlated with Factor 1 ($r = 0.34$, $p < 0.001$), total coping ($r = 0.27$, $p = 0.001$), and flexibility ($r = 0.17$, $p = 0.03$), but and was not significantly correlated with Factor 2 ($r = 0.15$, $p = 0.07$). Regarding divergent validity, some PACT factors correlated negatively with psychological distress as measured by HADS (Factor 1: $r = -0.38$, $p < 0.001$; total coping: $r = -0.25$, $p = 0.002$; flexibility: $r = -0.19$, $p = 0.01$) whereas PACT Factor 2 was not significantly correlated with distress ($r = -0.09$, $p = 0.28$).

Trauma Focus, Forward Focus, Total Coping, and Coping Flexibility Among Women With Breast Cancer

Overall, PACT scores in our sample were $M = 65.49$; $SD = 11.17$ for Factor 1—*Forward focus*; $M = 43.58$; $SD = 7.18$ for Factor 2—Trauma Focus; $M = 10.94$; $SD = 1.57$ for Total coping

TABLE 2 | Individual PACT item summaries for the total sample.

Item	Statistics			Percentage of endorsement							
	M (SD)	Sk	Ku	1	2	3	4	5	6	7	Total
1	5.24 (1.51)	−0.97	0.64	3.5	2.9	4.7	15.9	21.2	30.6	21.2	100
2	5.58 (1.43)	−1.34	1.61	2.3	2.9	3.5	9.9	15.2	38.6	27.5	100
3	5.39 (1.39)	−0.85	0.43	1.2	3.5	3.5	17.0	20.5	31.0	23.4	100
4	5.23 (1.46)	−0.92	0.48	2.4	3.5	6.5	14.1	21.8	33.5	18.2	100
5	5.48 (1.29)	−1.08	1.29	1.2	2.3	5.3	7.6	27.5	34.5	21.6	100
6	5.09 (1.37)	−0.39	−0.42	0.6	3.0	8.9	21.3	24.3	24.9	17.2	100
7	5.62 (1.55)	−0.94	−0.03	1.2	2.9	8.2	11.2	15.3	19.4	41.8	100
8	5.55 (1.49)	−1.12	0.95	2.3	3.5	2.3	12.9	19.9	25.7	33.3	100
9	5.42 (1.18)	−0.85	0.99	0.6	1.8	3.5	12.9	28.2	35.9	17.1	100
10	5.54 (1.35)	−0.79	0.07	0.6	1.8	5.9	14.2	19.5	28.4	29.6	100
11	5.18 (1.46)	−0.72	−0.11	1.2	4.8	7.7	16.1	19.6	32.1	18.5	100
12	5.94 (1.11)	−1.11	1.04	0.0	1.2	1.8	8.8	15.9	35.3	37.1	100
13	5.36 (1.30)	−0.72	0.05	0.0	4.1	4.1	15.9	22.9	32.9	20.0	100
14	5.70 (1.21)	−0.84	0.16	0.0	1.2	4.8	10.1	20.8	32.7	30.4	100
15	5.44 (1.33)	−0.86	0.37	0.6	2.9	5.3	13.5	21.2	34.1	22.4	100
16	5.74 (1.24)	−1.03	0.86	0.6	0.6	5.3	8.8	19.4	32.9	32.4	100
17	5.85 (1.15)	−0.81	0.03	0.0	0.6	3.0	10.1	20.7	29.0	36.7	100
18	5.36 (1.19)	−0.67	0.49	0.6	1.2	4.1	16.6	27.2	33.7	16.6	100
19	5.47 (1.38)	−0.73	−0.05	0.6	2.4	5.9	15.4	20.1	27.2	28.4	100
20	5.00 (1.36)	−0.53	−0.00	1.2	4.1	7.1	21.2	27.6	25.3	13.5	100

For each item of the PACT scale, the mean, standard deviation, and the percentage of endorsement for each possible item score (range 1–7) is displayed. PACT, Perceived Ability to Cope with Trauma; M, Mean; SD, Standard deviation; Sk, Skewness; Ku, Kurtosis.

and M = 9.85; SD = 2.76 for Flexibility. No differences were found when comparing the two clinical samples (CT and ET) on the PACT scale subscores (**Table 4**), revealing similar coping strategies and flexibility irrespective of the breast cancer treatment plan.

DISCUSSION

The present study describes the cross-cultural translation and adaptation of the PACT scale into European Portuguese, aiming at evaluating its psychometric properties, in terms of dimensionality, reliability (internal consistency), and construct (convergent and divergent) validity, in a sample of patients with breast cancer. Our results supported a two-factor structure for the PACT scale, consistent with the forward focus and trauma focus subscales proposed by the original authors (Bonanno et al., 2011). Noteworthy, the factorial structure of the original PACT scale was demonstrated in assessments of a potentially highly trauma-exposed Israeli sample and a group of American college students, in Hebrew and English, respectively (Bonanno et al., 2011). The other known published factorial structure of the PACT refers to its Italian version (Saita et al., 2017). The purpose of Saita et al. (2017) was to examine the factorial structure of the PACT scale in an Italian sample that shared characteristics with the individuals involved in the original validation study, thus including college students not directly exposed to potentially traumatic events but that could potentially present high levels of distress. The final most appropriate factor structure of the Italian version resulted in a total of 14 items, instead of the original 20 (Saita et al., 2017). The authors explain this difference based on different cultural aspects, highlighting the challenges of cross-cultural measurements (Saita et al., 2017).

Our study investigated the psychometric properties of the PACT scale in a different and novel sample: women recently diagnosed with breast cancer, that also constitutes a potentially traumatic event. Even though Portuguese shares a common Latin origin with Italian, representing two similar cultures, we were able to confirm the two-factor solution and preserve the 20-item scale used originally by Bonanno et al. (2011). In our results, item 10 ("Reduce my normal social obligations") presented a borderline factor loading, but we decided to retain it as our model showed an overall very good fit. In the Italian version, item 10 was eliminated due to very low factor loading in an Exploratory Factor Analysis (Saita et al., 2017). We believe that the concept of "social obligations" may have different meanings between Hebrew/English and Portuguese/Italian: while in the first it would be related to social responsibilities, in the latter it could be interpreted as social events, leading participants to rate how they would be able to reduce their usual social life events following a potentially traumatic event. Methodologically, as was also pointed out by Saita et al. (2017), these studies highlight the challenges associated with adapting existing instruments to a different culture following the etic approach (Tran et al., 2018). The etic approach argues that psychological processes are universal in nature and that instruments developed in a specific population could be applied to another (Saita et al., 2017; Tran et al., 2018). In the process of cross-cultural adaptation of the

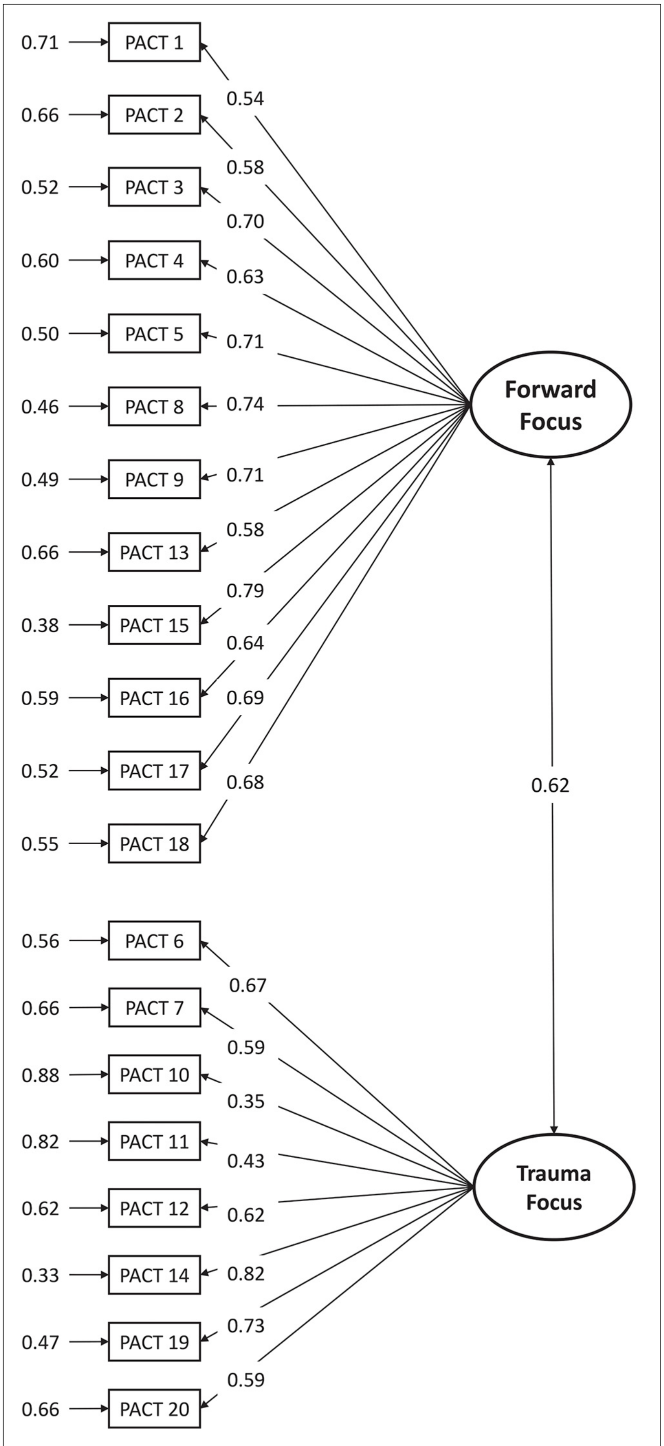

FIGURE 2 | Confirmatory factor analysis of the two-factor model of the PACT. Standardized coefficients and measurement errors are shown.

PACT scale, we assumed *a priori* that coping flexibility would be similar across the original Hebrew/English and Portuguese. Overall, our results confirmed both conceptual and statistical equivalences between the two versions.

Dimensionality findings were supported by the reliability results. The Portuguese version of the PACT was reliable when used to examine breast cancer patients. Factor 1 ("Forward focus" scale—12 items) showed an excellent reliability ($\omega = 0.91$; $\alpha = 0.90$), whereas a good value ($\omega = 0.82$; $\alpha = 0.83$) was found for Factor 2 ("Trauma focus" scale—8 items). The values remained stable with removal of any item on both factors, and the item-total correlations presented high levels, therefore confirming the theoretical structure of the two subscales. The two factors were moderately correlated. Similar results were displayed in the original version, where the 12 item-Forward focus also presented excellent reliability (0.91) and the 8 item-Trauma focus had a lower value (0.79) (Bonanno et al., 2011). Likewise, the Italian version of the PACT revealed a higher reliability of factor 1 (0.87) than of factor 2 (0.70) (Saita et al., 2017). Another study involving patients with breast cancer found similar results for forward focus (0.93) and trauma focus (0.74) (Hamama-Raz et al., 2012).

Concerning convergent and divergent validity, all PACT scores were related to high self-efficacy to coping with cancer (CBI-B), and most were related to high perceived QoL (QLQ-C30) and to low psychological distress (HADS). Our findings are consistent with previous research that demonstrated convergent validity with measures of positive cognitive–emotional regulation, ego resiliency, and optimism (Bonanno et al., 2011), thus suggesting that individuals with greater perceived coping ability and flexibility are more likely to experience positive emotions. On the other hand, as stated by the original authors, PACT scores are expected to show mild inverse associations with negative affect and with anxious or avoidant attachment (Bonanno et al., 2011), therefore confirming divergent validity. Our results confirmed that the PACT scores are positively associated with better adjustment to the diagnosis of cancer, and inversely related to psychological distress.

Furthermore, the two clinical samples (CT and ET) had similar results on all the PACT scale subscores, proving similar coping strategies and coping flexibility irrespective of the breast cancer treatment plan. The PACT and other psychosocial measures assessment were performed just after the diagnosis, and therefore either before the start of chemotherapy or within 2 weeks from the start of the endocrine therapy. Although patients were already aware of their therapeutic plan, no interpretations about coping strategies to deal with the trauma of the specific treatments and related side-effects can be made. The diagnosis of breast cancer is known to be particularly stressful, leading women to adapt coping strategies to address this challenge (Matthews et al., 2017).

Even though cancer is not an acute traumatic event, multiple studies have sought to measure the presence of PTSD in patients with cancer (Cordova et al., 2001, 2007, 2017; Mehnert and Koch, 2007; Foster et al., 2009; Arnaboldi et al., 2017; Matthews et al., 2017). It is important to highlight that, in the context of cancer, the threat is not only related to the present but also to the future, and patients are thus required to manage worry or distress about future health (Pat-Horenczyk et al., 2016). Experiencing positive emotions and reducing painful emotions, maintaining plans and goals, thinking optimistically, remaining calm, and being able to laugh, as well as processing the traumatic event,

TABLE 3 | Convergent and divergent validities.

	PACT Factor 1 *(forward focus)*	PACT Factor 2 *(trauma focus)*	PACT total coping	PACT flexibility
		Convergent validity		
CBI-B—total score	0.36***	0.32***	0.41***	0.26**
EORTC QLQ-C30 QoL	0.34***	0.15^{NS}	0.27**	0.17*
		Divergent validity		
HADS—total score	–0.38***	$–0.09^{NS}$	–0.25**	–0.20*

*PACT, Perceived Ability to Cope with Trauma; CBI-B, Cancer Behavior Inventory brief version; HADS, Hospital Anxiety and Depression Scale; EORTC QLQ-C30, European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire. ***p < 0.001; **p < 0.01; *p < 0.05; NS, non-significant.*

TABLE 4 | Performance on the PACT scale.

	Chemotherapy group	Endocrine therapy group	*t*	*p*
PACT Factor 1—forward focus, mean (SD)	64.64 (11.79)	66.64 (10.25)	−1.14	0.26
PACT Factor 2—trauma focus, mean (SD)	43.14 (7.45)	44.16 (6.81)	−0.89	0.37
PACT total coping, mean (SD)	10.80 (1.61)	11.11 (1.51)	−1.22	0.22
PACT flexibility, mean (SD)	9.56 (2.95)	10.25 (2.44)	−1.61	0.11

Comparisons between groups were carried out by independent sample t-tests.

allows individuals to evoke powerful changes in their emotional trajectory (Bonanno, 2005, 2004; Bonanno et al., 2011). In this regard, the PACT scale has been shown to be useful in measuring coping flexibility and in moderating the impact of heightened trauma exposure to breast cancer (Hamama-Raz et al., 2012; Pat-Horenczyk et al., 2016; Brivio et al., 2021).

Hamama-Raz et al. (2012) aimed to explore the causes affecting the decision of patients with breast cancer to participate in group intervention, after the end of adjuvant therapy, based on an approach to enhance resilience. Significantly higher levels of coping flexibility on PACT were reported by women who did not show an interest in the group intervention, relative to those who participated (Hamama-Raz et al., 2012). The authors suggest that women employing appropriate and psychologically healthy ways of coping are more likely to reveal resilience and emotional regulation, and may therefore perceive group therapy as not necessary or redundant (Hamama-Raz et al., 2012). In another post-cancer treatment study, Pat-Horenczyk et al. (2016) employed the PACT scale to assess the level of coping flexibility of female breast cancer patients over a 2-year period. The objectives of their study were to identify different post-treatment adaptation profiles, factors that predicted the adaptation profiles, and trajectories and transitions in the adaptation profiles. Four post-cancer treatment adaptation profiles were suggested by the authors: distressed, resistant, constructive growth, and struggling growth (Pat-Horenczyk et al., 2016). In the specific case of coping flexibility, it significantly predicted the likelihood of belonging to a particular profile, i.e., higher levels of flexibility increased the odds of being classified as either struggling or constructive growths, as well as decreased the likelihood of belonging to the distressed profile (Pat-Horenczyk et al., 2016).

In a very recent study, Brivio et al. (2021) showed that coping flexibility contributed significantly to manage the positive and negative affect in patients with cancer during the COVID19 pandemic in Italy. The positive states were evidenced by the PACT Trauma focus subscale, confirming that the perceived ability to focus on processing the trauma is associated with positive states (Brivio et al., 2021). Altogether, these results proved that higher levels of coping flexibility enhances resilience and emotional regulation (Hamama-Raz et al., 2012), are associated both post-traumatic growth and distress (Pat-Horenczyk et al., 2016), and helps activate a more positive outlook and think realistically about COVID-19 (Brivio et al., 2021), among breast cancer patients. Noteworthy, the available PACT versions are limited to English, Hebrew, and Italian, with studies performed in patients with breast cancer performed with the latter two. Our methodological approach, including analysis of the PACT scale validity; reliability; and responsiveness, correspond to the recommended quality criteria for measurement properties of health status questionnaires (Terwee et al., 2007) and of assessment tools in cancer patients (Tian et al., 2019).

In the present study we assessed women with early breast cancer just before starting treatment, aiming at evaluating the potentially traumatic experience of receiving a breast cancer diagnosis. Our sample was subdivided according to treatment plan (chemo vs. endocrine therapy). Some limitations should consequently be noted, mainly related to psychometric assessments that could not be performed. First, the sample size represents a limitation that prevented the performance of a differential item functioning analysis or a multigroup CFA. These analyses would enable us to show the extent to which an item might be measuring different abilities between members of different groups or to psychometrically determine whether the PACT would elicit a similar response pattern across our two subsamples. Secondly, test-retest reliability could not be performed as a retest assessment would probably lead to a different condition while answering to the PACT questions, as patients would be at a treatment phase and coping would most likely be related to treatment and not to diagnosis. We therefore consider that assessing test-retest reliability could lead to misinterpretations and opted not to include it. Future studies should include patients with other types of cancer or different cancer stages, or samples related to other traumatic events, thus providing additional information about the PACT and its applicability for Portuguese patients.

Despite these limitations, the current study contributes to the literature about cross-cultural adaptation and measurement by examining, for the first time, the psychometric properties of the PACT scale in a Portuguese sample of women with breast cancer. Our results indicate that the original two-factor structure is applicable in the current sample. Moreover, a similar organization of coping styles is maintained, and a latent measure of coping flexibility can be calculated. Using this measure in clinical practice may contribute to understand how patients cope with potentially traumatic events, thus helping to provide proper interventions to achieve better psychological adjustments.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Champalimaud Foundation. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RL and AO-M conceived and designed the work. RL and DF were responsible for the translation process of the PACT scale. DF and BS acquired the data, including assessment of eligibility criteria of patients participating in the study. RL, BC, SA, DF, and AO-M analyzed and interpreted data. BC and DF extracted clinical and demographic data with input from RL, BS, and AO-M. RL, BC, and AO-M drafted the manuscript, which was critically revised by the remaining authors for important intellectual content. AO-M supervised the research and acts as corresponding author. All authors contributed to the article and approved the submitted version.

FUNDING

RL was supported by the 2018 Scientific Employment Stimulus from Fundação para a Ciência e Tecnologia, Portugal (CEECIND/04157/2018). DF, BS, and AO-M were supported by the BOUNCE project (grant agreement no. 777167), and AO-M was supported by the FAITH project (grant agreement no. 875358), both funded by the European Union's Horizon 2020 Research and Innovation Programme.

ACKNOWLEDGMENTS

We would like to thank clinicians from the Breast and Neuropsychiatry Units of the Champalimaud Clinical Centre, for assistance in patient recruitment and assessment under the BOUNCE project, and BOUNCE Consortium members for contributing to the selection of the psychological instruments for the BOUNCE clinical study.

REFERENCES

Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., et al. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *JNCI J. Natl. Cancer Instit.* 85, 365–376. doi: 10.1093/jnci/85.5.365

Abbey, G., Thompson, S. B. N., Hickish, T., and Heathcote, D. (2015). A meta-analysis of prevalence rates and moderating factors for cancer-related post-traumatic stress disorder. *Psycho Oncol.* 24, 371–381. doi: 10.1002/pon.3654

Adams, R. N., Mosher, C. E., Cohee, A. A., Stump, T. E., Monahan, P. O., Sledge, G. W. Jr., et al. (2017). Avoidant coping and self-efficacy mediate relationships between perceived social constraints and symptoms among long-term breast cancer survivors. *Psycho Oncol.* 26, 982–990. doi: 10.1002/pon.4119

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*, 4th Edn. Washington, DC: American Psychiatric Association.

Arnaboldi, P., Riva, S., Crico, C., and Pravettoni, G. (2017). A systematic literature review exploring the prevalence of post-traumatic stress disorder and the role played by stress and traumatic stress in breast cancer diagnosis and trajectory. *Breast Cancer* 9, 473–485. doi: 10.2147/BCTT.S111101

Bartholomew, T. T., Brack, A. S. B., Leak, G. K., Hearley, A. R., and McDermott, T. J. (2017). Perceived ability to cope with trauma among U.S. combat veterans. *Mil. Psychol.* 29, 165–176. doi: 10.1037/mil0000150

Basińska, M. A. (2015). "Polish version of coping flexibility scale–the summary of research results," in *Coping Flexibility with Stress in Health and in Disease. Elastyczne Radzenie Sobie Ze Stresem w Zdrowiu Iw Chorobie*, ed. A. B. Małgorzata (Bydgoszcz: Wydawnictwo UKW), 272–294.

Basińska, M. A., Kruczek, A., Borzyszkowska, A., Góralska, K., Grzankowska, I., and Sołtys, M. (2021). Flexibility in coping with stress questionnaire: structure and psychometric properties. *Curr. Issues Pers. Psychol.* 9, 179–194. doi: 10.5114/cipp.2021.106412

Bonanno, G. A. (2004). Loss, trauma, and human resilience: have we underestimated the human capacity to thrive after extremely aversive events? *Am. Psychol.* 59, 20–28. doi: 10.1037/0003-066X.59.1.20

Bonanno, G. A. (2005). Resilience in the face of potential trauma. *Curr. Direct. Psychol. Sci.* 14, 135–138. doi: 10.1111/j.0963-7214.2005.00347.x

Bonanno, G. A., Horenczyk, R. P., and Noll, J. (2011). Coping flexibility and trauma: the perceived ability to cope with trauma (PACT) scale. *Psychol. Trauma* 3, 117–129. doi: 10.1037/a0020921

Brandão, T., Schulz, M. S., and Matos, P. M. (2017). Psychological adjustment after breast cancer: a systematic review of longitudinal studies. *Psycho Oncol.* 26, 917–926. doi: 10.1002/pon.4230

Brandtstädter, J., and Renner, G. (1990). Tenacious goal pursuit and flexible goal adjustment: explication and age-related analysis of assimilative and accommodative strategies of coping. *Psychol. Aging* 5, 58–67. doi: 10.1037/0882-7974.5.1.58

Brivio, E., Guiddi, P., Scotto, L., Giudice, A. V., Pettini, G., Busacchio, D., et al. (2021). Patients living with breast cancer during the coronavirus pandemic: the role of family resilience, coping flexibility, and locus of control on affective responses. *Front. Psychol.* 11:567230. doi: 10.3389/fpsyg.2020.567230

Campbell-Enns, H., and Woodgate, R. (2015). The psychosocial experiences of women with breast cancer across the lifespan: a systematic review protocol. *JBI Database Syst. Rev. Implement. Rep.* 13, 112–121. doi: 10.11124/jbisrir-2015-1795

Cheng, C. (2001). Assessing coping flexibility in real-life and laboratory settings: a multimethod approach. *J. Pers. Soc. Psychol.* 80, 814–833. doi: 10.1037/0022-3514.80.5.814

Cheng, C., Lau, H. P. B., and Chan, M. P. S. (2014). Coping flexibility and psychological adjustment to stressful life changes: a meta-analytic review. *Psychol. Bull.* 140, 1582–1607. doi: 10.1037/a0037913

Cordova, M. J., Cunningham, L. L. C., Carlson, C. R., and Andrykowski, M. A. (2001). Posttraumatic growth following breast cancer: a controlled comparison study. *Health Psychol.* 20, 176–185. doi: 10.1037/0278-6133.20.3.176

Cordova, M. J., Giese-Davis, J., Golant, M., Kronenwetter, C., Chang, V., and Spiegel, D. (2007). Breast cancer as trauma: posttraumatic stress and posttraumatic growth. *J. Clin. Psychol. Med. Settings* 14, 308–319. doi: 10.1007/s10880-007-9083-6

Cordova, M. J., Riba, M. B., and Spiegel, D. (2017). Post-traumatic stress disorder and cancer. *Lancet Psychiatry* 4, 330–338. doi: 10.1016/S2215-0366(17)30014-7

Foster, C., Wright, D., Hill, H., Hopkinson, J., and Roffe, L. (2009). Psychosocial implications of living 5 years or more following a cancer diagnosis: a systematic review of the research evidence. *Eur. J. Cancer Care* 18, 223–247. doi: 10.1111/j.1365-2354.2008.01001.x

Galatzer-Levy, I. R., Burton, C. L., and Bonanno, G. A. (2012). Coping flexibility, potentially traumatic life events, and resilience: a prospective study of college student adjustment. *J. Soc. Clin. Psychol.* 31, 542–567. doi: 10.1521/jscp.2012.31.6.542

Gana, K., and Broc, G. (2018). *Structural Equation Modeling with Lavaan*. Hoboken, NJ: John Wiley & Sons, Inc. doi: 10.1002/9781119579038

Gr egoire, J. (2018). ITC guidelines for translating and adapting tests (second edition). *Int. J. Test.* 18, 101–134. doi: 10.1080/15305058.2017.1398166

Hack, T. F., Pickles, T., Ruether, J. D., Weir, L., Bultz, B. D., Mackey, J., et al. (2010). Predictors of distress and quality of life in patients undergoing cancer therapy: impact of treatment type and decisional role. *Psycho Oncol.* 19, 606–616. doi: 10.1002/pon.1590

Hair, J. F. (2011). "Multivariate data analysis: an overview," in *International Encyclopedia of Statistical Science*, ed. M. Lovric (Berlin: Springer), 904–907. doi: 10.1007/978-3-642-04898-2_395

Hamama-Raz, Y., Perry, S., Pat-Horenczyk, R., Bar-Levav, R., and Stemmer, S. M. (2012). Factors affecting participation in group intervention in patients after adjuvant treatment for early-study breast cancer. *Acta Oncol.* 51, 208–214. doi: 10.3109/0284186X.2011.648339

Heitzmann, C. A., Merluzzi, T. V., Jean-Pierre, P., Roscoe, J. A., Kirsh, K. L., and Passik, S. D. (2011). Assessing self-efficacy for coping with cancer: development and psychometric analysis of the brief version of the cancer behavior inventory (CBI-B). *Psycho Oncol.* 20, 302–312. doi: 10.1002/pon.1735

Henselmans, I., Fleer, J., van Sonderen, E., Smink, A., Sanderman, R. T., and Ranchor, A. V. (2011). The tenacious goal pursuit and flexible goal adjustment scales: a validation study. *Psychol. Aging* 26, 174–180. doi: 10.1037/a0021536

Jöreskog, K. G., and Sörbom, D. (1989). *LISREL 7: A Guide to the Program and Applications*. Chicago, IL: Scientific Software International.

Kant, J., Czisch, A., Schott, S., Siewerdt-Werner, D., Birkenfeld, F., and Keller, M. (2018). Identifying and predicting distinct distress trajectories following a breast cancer diagnosis - from treatment into early survival. *J. Psychosom. Res.* 115, 6–13. doi: 10.1016/j.jpsychores.2018.09.012

Karademas, E. C., Simos, P., Pat-Horenczyk, R., Roziner, I., Mazzocco, K., Sousa, B., et al. (2021). Cognitive, emotional, and behavioral mediators of the impact of coping self-efficacy on adaptation to breast cancer: an international prospective study. *Psycho Oncol.* 30, 1555–1562. doi: 10.1002/pon.5730

Kato, T. (2012). Development of the coping flexibility scale: evidence for the coping flexibility hypothesis. *J. Counsel. Psychol.* 59, 262–273. doi: 10.1037/a0027770

Knowles, L. M., and O'Connor, M. F. (2015). Coping flexibility, forward focus and trauma focus in older widows and widowers. *Bereav. Care* 34, 17–23. doi: 10.1080/02682621.2015.1028200

Kochaki Nejad, Z., Aghdam, A. M., Hassankhani, H., Jafarabadi, M. A., and Sanaat, Z. (2015). Cancer-related self-efficacy in Iranian women with breast cancer. *Womens Health Bull.* 2:e23248. doi: 10.17795/whb-23248

Lashbrook, M. P., Valery, P. C., Knott, V., Kirshbaum, M. N., and Bernardes, C. M. (2018). Coping strategies used by breast, prostate, and colorectal cancer survivors. *Cancer Nurs.* 41, E23–E39. doi: 10.1097/NCC.0000000000000528

Lazarus, R. S., and Folkman, S. (1984). *Stress, Appraisal, and Coping*. New York, NY: Springer Publishing Company.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* 48, 936–949. doi: 10.3758/s13428-015-0619-7

Linley, P. A., and Joseph, S. (2004). Positive change following trauma and adversity: a review. *J. Trauma. Stress* 17, 11–21. doi: 10.1023/B:JOTS.0000014671.27856.7e

Macía, P., Barranco, M., Gorbeña, S., and Iraurgi, I. (2020). Expression of resilience, coping and quality of life in people with cancer. *PLoS One* 15:e0236572. doi: 10.1371/journal.pone.0236572

Matthews, H., Grunfeld, E. A., and Turner, A. (2017). The efficacy of interventions to improve psychosocial outcomes following surgical treatment for breast cancer: a systematic review and meta-analysis. *Psycho Oncol.* 26, 593–607. doi: 10.1002/pon.4199

Mehnert, A., and Koch, U. (2007). Prevalence of acute and post-traumatic stress disorder and comorbid mental disorders in breast cancer patients during primary cancer care: a prospective study. *Psycho Oncol.* 16, 181–188. doi: 10.1002/pon.1057

Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*. New York, NY: McGraw-Hill.

Pais-Ribeiro, J., Pinto, C., and Santos, C. (2008). Validation study of the portuguese version of the QLC-C30-V.3. *Psicol. Saúde Doenças* 9, 89–102.

Pais-Ribeiro, J., Silva, I., Ferreira, T., Martins, A., Meneses, R., and Baltar, M. (2007). Validation study of a portuguese version of the hospital anxiety and depression scale. *Psychol. Health Med.* 12, 225–237. doi: 10.1080/13548500500524088

Park, M., Chang, E. R., and You, S. (2015). Protective role of coping flexibility in PTSD and depressive symptoms following trauma. *Pers. Individ. Differ.* 82, 102–106. doi: 10.1016/j.paid.2015.03.007

Pat-Horenczyk, R., Saltzman, L. Y., Hamama-Raz, Y., Perry, S., Ziv, Y., Frolich, R. G., et al. (2016). Stability and transitions in posttraumatic growth trajectories among cancer patients: LCA and LTA analyses. *Psychol. Trauma Theory Res. Pract. Policy* 8, 541–549. doi: 10.1037/tra0000094

Pereira, M., Izdebski, P., and Graça Pereira, M. (2021). Validation of the brief version of the cancer behavior inventory in breast cancer portuguese patients. *J. Clin. Psychol. Med. Settings* 28, 491–502. doi: 10.1007/s10880-021-09773-5

Pinciotti, C. M., Seligowski, A. V., and Orcutt, H. K. (2017). Psychometric properties of the PACT scale and relations with symptoms of PTSD. *Psychol. Trauma Theory Res. Pract. Policy* 9, 362–369. doi: 10.1037/tra0000206

Saita, E., Acquati, C., Fenaroli, V., Zuliani, C., and Bonanno, G. A. (2017). A confirmatory factor analysis of the perceived ability to cope with trauma (PACT) scale. *TPM Test. Psychom. Methodol. Appl. Psychol.* 24, 255–268. doi: 10.4473/TPM24.2.5

Shim, E.-J., Lee, J. W., and Min, Y. H. (2018). Does depression decrease the moderating effect of self-efficacy in the relationship between illness perception and fear of progression in breast cancer? *Psycho Oncol.* 27, 539–547. doi: 10.1002/pon.4532

Sullivan, K. A., and Wade, C. (2019). Assault-related mild traumatic brain injury, expectations of injury outcome, and the effect of different perpetrators: a vignette study. *Appl. Neuropsychol. Adult* 26, 58–64. doi: 10.1080/23279095.2017.1359603

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi: 10.3322/caac.21660

Swartzman, S., Booth, J. N., Munro, A., and Sani, F. (2017). Posttraumatic stress disorder after cancer diagnosis in adults: a meta-analysis. *Depress. Anxiety* 34, 327–339. doi: 10.1002/da.22542

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *J. Clin. Epidemiol.* 60, 34–42. doi: 10.1016/j.jclinepi.2006.03.012

Tian, L., Cao, X., and Feng, X. (2019). Evaluation of psychometric properties of needs assessment tools in cancer patients: a systematic literature review. *PLoS One* 14:e0210242. doi: 10.1371/journal.pone.0210242

Tran, T., Nguyen, T., and Chan, K. (2018). *Developing Cross-Cultural Measurement in Social Work Research and Evaluation*, Vol. 1. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780190496470.001.0001

Vriezekolk, J. E., van Lankveld, W. G. J. M., Eijsbouts, A. M. M., van Helmond, T., Geenen, R., and van den Ende, C. H. M. (2012). The coping flexibility questionnaire: development and initial validation in patients with chronic rheumatic diseases. *Rheumatol. Int.* 32, 2383–2391. doi: 10.1007/s00296-011-1975-y

Zhou, Y., MacGeorge, E. L., and Myrick, J. G. (2020). Mental health and its predictors during the early months of the COVID-19 pandemic experience in the United States. *Int. J. Environ. Res. Public Health* 17:6315. doi: 10.3390/ijerph17176315

Zigmond, A. S., and Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatr. Scand.* 67, 361–370. doi: 10.1111/j.1600-0447.1983.tb09716.x

Conflict of Interest: AO-M was national coordinator for Portugal of a non-interventional study (EDMS-ERI-143085581, 4.0) to characterize a Treatment-Resistant Depression Cohort in Europe, sponsored by Janssen-Cilag, Ltd. (2019–2020), is recipient of a grant from Schuhfried GmbH for norming and validation of cognitive tests, and is national coordinator for Portugal of trials of psilocybin therapy for treatment-resistant depression, sponsored by Compass Pathways, Ltd. (EudraCT number 2017-003288-36 and 2020-001348-25), and of esketamine for treatment-resistant depression, sponsored by Janssen-Cilag, Ltd. (EudraCT NUMBER: 2019-002992-33).

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 March 2022
doi: 10.3389/fpsyg.2022.792805

Psychometric Properties of the Zarit Burden Interview in Informal Caregivers of Persons With Intellectual Disabilities

Alicia Boluarte-Carbajal[1], Rubí Paredes-Angeles[2,3] and Arnold Alejandro Tafur-Mendoza[2,4]*

[1] Facultad de Ciencias de la Salud, Universidad César Vallejo, Lima, Peru, [2] Grupo de Estudios Avances en Medición Psicológica, Universidad Nacional Mayor de San Marcos, Lima, Peru, [3] Instituto Peruano de Orientación Psicológica, Lima, Peru, [4] Research Center (CIUP), Universidad del Pacífico, Lima, Peru

Intellectual disability leads to a loss of autonomy and a high level of dependence, requiring support from another person permanently. Therefore, it is necessary to incorporate the assessment of caregiver burden in healthcare actions, to avoid putting the health of caregivers and patients at risk. In this sense, the study aimed to analyze the internal structure of the Zarit Burden Interview (ZBI) in a sample of caregivers of people with intellectual disabilities, to provide convergent and discriminant evidence with a measure of the risk of maltreatment, and to estimate the reliability of the scores from the Classical Test Theory and the Rasch Measurement Theory. The study was instrumental. The sample consisted of 287 Peruvian informal primary caregivers of persons diagnosed with intellectual disabilities. To collect validity evidence, the internal structure (confirmatory factor analysis, CFA) and the relationship with other variables (convergent and discriminant evidence) were used, while reliability was estimated through the omega coefficient and Rasch analysis. The internal structure of the ZBI corroborated a unidimensional structure. In terms of convergent and discriminant evidence, the scale presents adequate evidence. Reliability levels were also good. Previously, the psychometric properties of the ZBI have not been studied in caregivers of people with intellectual disabilities, and it represents the first study of the scale in Peru. The results obtained will allow the use of this scale to design actions in the work with caregivers and studies to understand the psychology of the caregiver.

Keywords: Zarit Burden Interview, ZBI, caregivers, intellectual disabilities, Rasch analysis, psychometric properties

OPEN ACCESS

Edited by:
Paola Gremigni,
University of Bologna, Italy

Reviewed by:
Maribel Peró-Cebollero,
University of Barcelona, Spain
Zi-Wei Liu,
Hunan Normal University, China

****Correspondence:***
Alicia Boluarte-Carbajal
aliciabolucar@gmail.com

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *11 October 2021*
Accepted: *28 January 2022*
Published: *09 March 2022*

Citation:
Boluarte-Carbajal A, Paredes-Angeles R and Tafur-Mendoza AA (2022) Psychometric Properties of the Zarit Burden Interview in Informal Caregivers of Persons With Intellectual Disabilities. Front. Psychol. 13:792805. doi: 10.3389/fpsyg.2022.792805

INTRODUCTION

The concept of disability has evolved historically, before linked to a bio-medical model of disease and now based on a systemic model, and can be conceptualized as a health condition determined by the social context (World Health Organization [WHO], 2007). In this sense, the degree of disability depends on the interaction that the person establishes with others, being many times the attitudes of the environment an obstacle to personal, family, educational, social, or labor participation (Cuenot, 2018). It is estimated that 15.6% of people over 15 years of age live with a disability, of which 5.1% are under 14 years of age and 0.7% have a severe disability (World Health Organization [WHO] and World Bank, 2011). Disability in Peru represents 10.3% of the total population, of which, 48.3%

present visual disability, 7.6% hearing, 3.1% to speak or communicate, 15.1% to move or walk, 4.2% to understand or learn, 3.3% to relate to others, and those with multi disability represent 18.4% (Instituto Nacional de Estadística e Informática [INEI], 2019). However, there is no exact statistics on disability due to intellectual deficit.

Intellectual disability is one of the disabilities that generates greater dependence on its environment and is commonly associated with other comorbidities, such as cerebral palsy, autism, seizures, and sensory problems (Uzun et al., 2020). According to Schalock et al. (2021), it is defined as a deficit in intellectual functioning, which is expressed through an intelligence quotient (IQ) below 70, in addition to limitations in social adaptation skills, such as self-direction, social skills, movement, and personal care among others, presenting before the age of 18 years (Tassé et al., 2016). This type of disability leads to a loss of autonomy and a high level of dependence, requiring support from another person permanently, commonly referred to as a caregiver (Seguí et al., 2008).

According to the literature, there are two types of caregivers: formal and informal. The formal caregiver is the person who receives payment for the service. While the informal caregiver is commonly a family member, friend, or neighbor. Although this type of caregiver is neither paid nor trained to perform the work, they have a high degree of commitment characterized by their affective bond (Ruiz and Nava, 2012; Montero et al., 2014).

In the case of people with intellectual disabilities, the type of care in healthcare is usually informal, with the mother being the main caregiver (Boluarte, 2019). Although currently there is greater economic and social participation of women. However, still retains the traditional role in child-rearing (Duarte and García-Horta, 2016; Tartaglini et al., 2020). In turn, other family members also assume the role of caregiver, such as grandparents, siblings, and in some cases the father, who, in light of the existing literature, assumes little responsibility for the child-rearing (Espín, 2008; Lima-Rodríguez et al., 2018; Rodríguez et al., 2019). In this context, the presence of dysfunctionality in family relationships, inadequate coping styles, job abandonment, and low educational level of the mother not only affects the quality of family life (Espín, 2008; Cohen et al., 2014) but also the deterioration of physical, mental, and social health with serious psychopathological implications (Seguí et al., 2008).

In this way, caregiver burden has been studied, defined as the attitudes and emotional-affective reactions from the experiences of the role with repercussions in the personal, family, and social spheres (Zarit et al., 1980). Therefore, it is necessary to incorporate the assessment of caregiver burden in healthcare actions, to avoid putting the health of caregivers and patients at risk (Zarit et al., 1986; Schreiner et al., 2006).

There are different scales to measure caregiver burden (Crespo and Rivas, 2015), such as the Caregiver Burden Inventory (CBI; Novak and Guest, 1989), widely used in a different diagnostic groups, in children, adolescents, and adults with spinal cord injury (Farmer et al., 2018; Conti et al., 2019; Ortiz-Rubio et al., 2021). The CBI has been recently adapted to Spanish (Vázquez et al., 2019) obtaining a reduced version of 15 items. Likewise, the Screen for Caregiver Burden (SCB; Vitaliano et al., 1991), is used for the objective and subjective measurements of caregiver burden in the elderly.

One of the most widely used instruments is the Zarit Burden Interview (ZBI; Zarit et al., 1980). This measure, unidimensional and with 29 items in its original version, assesses the caregiver's health, psychological wellbeing, finances, social life, and the relationship between the caregiver and the person with a disability. However, its factor structure and extent have had modifications over time (Knight et al., 2000; Yu et al., 2018).

The 22-item version (Zarit et al., 1985) has been studied in different contexts and translated into different languages, such as the United Kingdom (Siegert et al., 2010), Turkey (Özlü et al., 2009), Singapore (Seng et al., 2010), Portugal (da Cruz, 2010), Brazil (Taub et al., 2004), and among others. The adaptation to Spanish was carried out by Martín-Carrasco et al. (1996) and later re-evaluated by Martín-Carrasco et al. (2010), who showed that this measure was made up of three factors (i.e., burden, competence, and dependence) showing correlation with the caregiver's mental health status and the presence of behavioral disorders in the patient.

In its factorial structure, it has been found the presence of three factors with different denomination: impact of care, interpersonal burden, and self-efficacy expectations (Montorio et al., 1998); shame, anger, and self-criticism (Knight et al., 2000); tensions referring to the role, intrapsychic tensions, competencies, and expectations (Bianchi et al., 2016); subjective impact, competence, and dependence (Martín-Carrasco et al., 2010); role-related strain, self-criticism, and negative emotion (Tang et al., 2016). Other studies report the presence of five factors: sacrifice, loss of control, shame/anger, self-criticism, and dependence (Lu et al., 2009), caregiver feeling of over-sacrifice, dependency, negative emotion, caregiver feeling of inadequacy, and uncertainty about the patient's future (Ko et al., 2008). Other studies report four-factor structures: personal effort, privacy conflict, uncertain attitude, and guilt (Yoon and Robinson, 2005); consequences of caregiving on the caregiver, patient dependency, caregiving fatigue, uncertainty, guilt, and fear (Al-Rawashdeh et al., 2016); interpersonal burden, the impact of caregiving, competencies, and expectations about caregiving (Barreto-Osorio et al., 2015). The different theoretical interpretation in the measurement of caregiver burden with the ZBI is evident, demonstrating that there is no clear consensus on the dimensionality of the instrument (Monreal and Prieto, 2017).

Short versions are proposed to achieve a parsimonious structural model, which allows the instrument to be used in interventions that require optimizing the application time. In this sense, Hébert et al. (2000) with a 12-item version proposes a two-factor model: personal strain and role strain. Likewise, Bédard et al. (2001) proposed a short version of 12 and a 4-item version for screening studies, reporting its usefulness in caregivers of patients with different diagnoses (caregivers of older adults and children with chronic disabilities). The 12-item unidimensional version has shown adequate psychometric properties in several studies (O'Rourke and Tuokko, 2003; Ballesteros et al., 2012; Lin et al., 2017; Rueda et al., 2017; Pinyopornpanish et al., 2020; Tartaglini et al., 2020).

The literature shows that there is no clear consensus on the dimensionality of the instrument (Monreal and Prieto, 2017).

While some studies state the presence of three dimensions, others report the presence of one or more additional factors. Li et al. (2018) named the fourth-factor as caregiving performance corresponding to items 20 and 21. However, Barreto-Osorio et al. (2015) indicate that these items would be related to indecisiveness about caregiving (7, 20, and 21). The conceptual controversies of the construct could be explained by a solid theoretical basis. The updated theory of attachment and emotional self-regulation allows understanding the conflicts in close relationships (Jarrett, 1985; Mikulincer and Shaver, 2005), caring for people dependent on the family context, generates positive and negative affective relationships, which alter the quality of the relationship with repercussions on the quality of family life.

Reliability has been commonly measured by internal consistency, in the case of global, measures the coefficients fluctuated between 0.70 and 0.93 (Lu et al., 2009; Özlü et al., 2009; Bianchi et al., 2016). While in multidimensional analysis, coefficients below 0.70 are reported (Hébert et al., 2000). Other methods, such as test-retest and inter-observer reliability, have also been used, using the interclass correlation coefficient (ICC) technique, obtaining good results, e.g., ICC = 0.88 (Taub et al., 2004) and ICC = 0.78 (Rajabi-Mashhadi et al., 2015) in samples of caregivers of chronically disabled patients.

On the other hand, studies show the relationship of caregiver burden with other variables, such as the risk of caregiver abuse (Özcan et al., 2017; Orfila et al., 2018; Saravia et al., 2019). A positive correlation has also been observed (r = 0.844; p < 0.001) with abuse of caregivers of people with behavioral disorders affected by dementia to a moderate degree (Gimeno et al., 2021), indicating that the higher the level of burden, the risk of maltreatment increases by the caregiver. Likewise, in people with dementia in Spain, a moderate relationship (r = 0.486; p < 0.001) was found between the risk of caregiver abuse, measured through the Caregiver Abuse Screen (CASE), and caregiver burden, assessed by the ZBI (Rivera-Navarro et al., 2018).

Most of the ZBI psychometric studies have been tested in samples of caregivers of adult patients with mental and degenerative diseases (e.g., dementia, Alzheimer's, and acquired brain injury). To date, there are no psychometric studies that have assessed the ZBI in caregivers of people with intellectual disabilities. However, there are studies on the burden of caregivers in this population. Kim et al. (2021) found a partial mediation of family functioning on the relationship between care burden and quality of life for caregivers of children with intellectual disabilities in Mongolia. For the measurement of care burden, they used the CBI, reporting good levels of reliability (α > 0.70). In addition, Barros et al. (2019) found a negative relationship between quality of life domains (physical, psychological, social relationships, environment, and global) and the burden of caregivers of children and young adults with intellectual disabilities in Brazil. The assessment of burden of caregivers was conducted through the ZBI. However, due to the nature of the study, they only reported the overall estimate of reliability, which was good (α = 0.90).

The psychometric properties of the ZBI have not been examined in the Peruvian context, being relevant to know its psychometric properties to recommend its appropriate use in clinical-therapeutic assessment and intervention. Therefore, the present study is aimed to analyze the internal structure of the Zarit Burden Inventory in a sample of Peruvian primary caregivers of people with intellectual disabilities and to test the convergent and discriminant validity with a measure of the risk of maltreatment. In addition, we seek to estimate reliability using Classical Item Theory and Rasch Measurement Theory.

MATERIALS AND METHODS

Study Design

The study was instrumental in that it analyzed the psychometric properties of a scale that measures caregiver burden (Ato et al., 2013). The development of the study followed the guidelines proposed in the standards for educational and psychological tests (American Educational Research Association [AERA] et al., 2014). Complementarily, recommendations based on good practices for the development and review of scales in social, health, and behavioral sciences were considered (Boateng et al., 2018).

Participants

The selection of the participants was carried out through intentional non-probability sampling (Kerlinger and Lee, 2000). To determine the sample size, the recommendations for conducting a factor analysis were followed. In this sense, a ratio of three indicators per factor, a one-factor solution, low levels of communality, and an excellent-level criterion agreement (0.98), was considered *a priori*, obtaining a recommended sample size of 150 participants (Mundfrom et al., 2005). From this result, it was sought to obtain a sample size greater than the minimum recommended.

The initial sample size was 303. However, after the elimination of 16 outliers, the final study sample consisted of 287 informal primary caregivers of persons diagnosed with intellectual disabilities. The age of the caregivers ranged from 19 to 76 years (median = 40, median absolute deviation = 10.38). Regarding the characteristics of the caregivers, the majority were women (84.67%), the predominant marital status was married (40.21%), the highest proportion had a completed high school education (52.26%), had no illnesses (72.13%), were 24-h caregivers (45.64%), and most were fathers or mothers of the patients (86.76%). The persons with intellectual disabilities were aged between 1 and 64 years (median = 9, median absolute deviation = 4.45), and the highest percentage was moved without assistance (89.90%). A detailed description of the main demographic characteristics of the study participants is presented in **Table 1**.

Measures

Zarit Burden Interview

This instrument was designed by Zarit et al. (1980) to measure the perception of primary caregiver strain. In this study, the version adapted to Spanish by Rueda et al. (2017) was used in a group of Colombian family caregivers. The ZBI is composed

TABLE 1 | Sociodemographic characteristics of the participants (n = 287).

Variable	Category	Frequency	Percentage
Sex of caregiver	Male	44	15.33
	Female	243	84.67
Marital status	Single	52	18.18
	Married	115	40.21
	Cohabitant	92	32.17
	Separated	19	6.64
	Divorced	1	0.35
	Widower	7	2.45
Education level	Primary	25	8.71
	Secondary	150	52.26
	Technical superior	69	24.04
	Superior university	42	14.63
	No education	1	0.35
Caregiver with illness	Yes	80	27.87
	No	207	72.13
Relationship to patient	Parent	249	86.76
	Brother/sister	9	3.14
	Grandmother	14	4.88
	Uncle/aunt	7	2.44
	Another	8	2.79
Care hours	Between 1 and 5 h	21	7.32
	Between 6 and 10 h	43	14.98
	Between 11 and 15 h	28	9.76
	Between 16 and 20 h	29	10.10
	Between 21 and 24 h	35	12.20
	24 h a day	131	45.64
Main reason for caring	Own initiative	216	75.26
	Family decision	56	19.51
	Only one who could	15	5.23
Time as caregiver	Less than 1 year	17	5.92
	Between 1 and 3 years	32	11.15
	Between 3 and 6 years	62	21.60
	Between 6 and 9 years	62	21.60
	More than 10 years	114	39.72
Has another job	Yes	130	45.30
	No	157	54.70
Patient displacement	Moves with assistance	29	10.10
	Can move without assistance	258	89.90

of 22 items that are answered on a five-point Likert-type scale (Never = 0; Rarely = 1; Sometimes = 2; Quite often = 3; and Almost always = 4). The ZBI items assess the perceived impact of caregiving on the caregiver's physical health, emotional health, social activities, and financial situation. Overall ZBI scores range from 0 to 88 points, where a high score implies a greater perceived caregiver burden.

Caregiver Abuse Screen

This brief measure was designed by Reis and Nahmiash (1995) and is self-administered by caregivers. The CASE aims to detect the risk of physical, psychosocial abuse, and neglect by primary caregivers toward older adults (Reis and Nahmiash, 1995). For this study, the Spanish version of the CASE developed by Pérez-Rojo et al. (2015) was used. The CASE is made up of eight items grouped into two factors, Abuse (six items) and Neglect/Dependency (two items), which were found in the original study (Reis and Nahmiash, 1995) and the Spanish version (Pérez-Rojo et al., 2015). However, the Brazilian (Reichenheim et al., 2009) and Pakistani (Khan et al., 2020) versions suggest the presence of a unidimensional structure. The items are answered on a dichotomous response scale (No = 0; Yes = 1), so, their total scores vary between 0 and 8, where a higher score indicates a high risk of maltreatment.

In this study, for the collection of validity evidence based on internal structure, a two-factor related model was tested through confirmatory factor analysis (CFA), obtaining adequate fit indices [χ^2 = 41.284, df = 19, χ^2/df = 2.173, Comparative Fit Index (CFI) = 0.973, Tucker-Lewis Index (TLI) = 0.961, Root Mean Square Error of Approximation (RMSEA) = 0.064 [90% confidence interval (CI): 0.037, 0.091], Standardized Root Mean Square Residual (SRMR) = 0.070, and Weighted Root Mean Square Residual (WRMR) = 0.885]. However, scores on the Neglect/Dependency factor presented a low level of reliability (ω = 0.412), unlike the abuse factor (ω = 0.768), which obtained an acceptable level. Because of these results and the high correlation between the factors (r = 0.940), a unidimensional model was tested. The CFA indicated good fit indices for the unifactorial structure (χ^2 = 41.381, df = 20, χ^2/df = 2.069, CFI = 0.974, TLI = 0.964, RMSEA = 0.061 [90% CI: 0.034, 0.088], SRMR = 0.070, and WRMR = 0.889) and the factor loadings were found to be between 0.527 and 0.876. Likewise, the average variance explained (AVE) was 0.491, showing an acceptable level of convergent evidence. Regarding reliability, the CASE showed acceptable internal consistency (ω = 0.793 [95% CI: 0.733, 0.824]).

Procedure

People with intellectual disabilities from special basic education schools and private psychological centers that serve this population were identified, and their main caregivers were identified. The identified caregivers signed an informed consent form, where the objective of the research was explained to them and the guarantee of confidentiality and anonymity of their participation. Subsequently, the caregivers proceeded to fill out the data collection form in person. This form consisted of a sociodemographic survey and the two measurement instruments (ZBI and CASE).

Once the database with 303 responses was obtained, univariate outliers were analyzed through the median absolute deviation (Leys et al., 2019) and multivariate outliers through the Mahalanobis-MCD distance (Leys et al., 2018). Four cases of univariate outliers and 12 cases of multivariate outliers were found, being removed from the database, leaving the final database with 287 participants. This process was carried out because many statistical procedures are affected by the presence of outliers. In some situations, in the presence of outliers, the statistical power of some methods presents less power and therefore, unreliable results (Aguinis et al., 2013). Likewise, outlier detection is a recommended good practice in data management. Finally, a sensitivity analysis was performed with and without outliers, observing a small impact of outliers, mainly in descriptive statistics, such as the mean (Thabane et al., 2013).

Ethical Considerations

The present study was carried out with the commitment to ethical standards and values, where the researchers assume total responsibility and veracity demonstrating each result obtained, likewise, the reliability of the data has been acquired respecting the anonymity of the participants, in such a way that the personal information concerning those evaluated in the study is not known. The study was authorized by the ethics committee of the Universidad César Vallejo. Participants completed informed consent form to respond to the measurement instruments. All procedures performed in studies involving human participants were following the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Data Analysis

Statistical analyses were carried out in five stages. In the first stage, the descriptive measures of the items were obtained through the mean, standard deviation, skewness, and kurtosis. These last two coefficients indicated the level of departure from a normal distribution, considering adequate values between −2 and 2 (Tabachnick and Fidell, 2019). Likewise, the floor effect and ceiling effect of the items were analyzed, considering the percentage of people who answered the lowest and highest answer alternative, respectively. In this sense, those items with percentages equal to or less than 15% were evaluated as free of these effects (McHorney and Tarlov, 1995). Additionally, the discrimination of the items was estimated through the corrected item-rest polyserial correlation, taking as acceptable indices greater than 0.20 (Schmeiser and Welch, 2006).

In the second stage, validity evidence was collected based on the internal structure of the test using CFA. The estimation method was the Weighted Least Squares Means and Variance adjusted (WLSMV) with robust SEs and Scaling-Shifted scaled statistic test (SS), applied to the matrix of polychoric correlations of the items. To fix the metrics of the dimensions, one of their indicators, called the reference indicator, was used. That is, the factor loadings of the first indicator with its dimension were fixed. These fixed loadings were equal to 1. Regarding the goodness-of-fit indices to assess the estimated models, the ratio between Chi-square and degrees of freedom ($SS\chi^2/df$) was used, taking as appropriate values below 5 (Schumacker and Lomax, 2016); the CFI and TLI with adequate values higher than 0.90 (Keith, 2019); the RMSEA and SRMR considering values less than 0.08 as adequate (Schumacker and Lomax, 2016); and the WRMR with values lower than 1 being appropriate (DiStefano et al., 2018). Likewise, factor loadings above 0.40 were considered acceptable (Brown, 2015).

Later, in the third phase, validity evidence was collected based on the relationship with other variables. For this purpose, convergent and discriminant evidence was collected. The convergent evidence was evaluated from the average variance extracted (AVE), taking as minimum acceptable values those proposed by Moral (2019), which considers the factor loadings, the reliability coefficient, and the number of factor items evaluated. A structural equation model (SEM) was tested to estimate the relationship between ZBI and CASE, using the same criteria as in the CFA to assess model fit. In addition, the relationship between the variables was assessed as small, medium, and large considering correlation coefficients above 0.10, 0.30, and 0.50, respectively (Cohen, 1988). On the other hand, the discriminant evidence was collected through two procedures, the heterotrait-monotrait ratio (HTMT), taking as adequate values lower than 0.85 (Henseler et al., 2015), and the Fornell and Larcker criterion, which consists of comparing the square root of the AVE and the correlations with the other variables, where the former must be greater than the latter to conclude that there is discriminant evidence (Fornell and Larcker, 1981).

In the fourth phase of analysis, the reliability of the test scores was evaluated using the internal consistency method. For this objective, the ordinal omega coefficient was used, estimated from the factorial solution obtained from the CFA (Viladrich et al., 2017; Flora, 2020). This coefficient varies between 0 and 1, being valued as adequate from 0.70 (Nunnally and Bernstein, 1994). CIs were estimated at a 95% confidence level using the bias-corrected and accelerated bootstrap method with 10,000 replications. Additionally, to have a better understanding of the score reliability, inter-item polychoric correlations were estimated (Ventura-León and Peña-Calero, 2021).

TABLE 2 | Descriptive statistics, the proportion of responses, and discrimination of the items.

Item	*M*	SD	Sk	Ku	Item-rest correlation	Responses (%) 0	1	2	3	4
1	2.12	1.25	0.08	−0.91	0.43	10.80	19.86	36.59	12.20	20.56
2	1.56	1.36	0.39	−1.02	0.51	31.01	18.82	26.13	11.50	12.54
3	1.28	1.23	0.60	−0.58	0.69	36.93	19.86	28.57	7.67	6.97
4	0.48	0.94	2.08	3.81	0.53	73.52	12.20	9.41	2.44	2.44
5	0.39	0.80	2.24	4.88	0.51	76.31	12.54	8.36	1.74	1.05
6	0.55	0.97	1.65	1.76	0.54	70.38	11.15	12.54	4.53	1.39
7	2.20	1.43	−0.08	−1.29	0.44	15.68	17.42	27.18	10.45	29.27
8	2.40	1.35	−0.32	−1.06	0.36	11.85	13.59	26.48	18.47	29.62
9	0.66	1.05	1.53	1.52	0.62	64.81	13.94	14.63	3.48	3.14
10	0.64	1.02	1.47	1.24	0.59	66.20	12.20	15.33	4.18	2.09
11	0.72	1.03	1.38	1.25	0.57	58.54	19.86	15.33	3.48	2.79
12	0.71	1.10	1.45	1.12	0.63	63.76	13.94	13.59	5.23	3.48
13	0.55	0.93	1.63	1.93	0.55	68.29	13.94	13.24	3.14	1.39
14	1.97	1.50	0.06	−1.41	0.40	23.69	17.77	20.56	13.59	24.39
15	1.82	1.40	0.22	−1.18	0.56	23.00	20.91	25.44	12.20	18.47
16	0.91	1.11	1.06	0.28	0.29	49.48	23.00	18.12	5.92	3.48
17	0.67	1.02	1.42	1.07	0.62	63.07	16.38	13.24	5.57	1.74
18	0.62	0.99	1.50	1.45	0.25	66.20	12.54	16.38	2.79	2.09
19	1.01	1.17	0.89	−0.15	0.39	47.74	18.47	23.34	5.92	4.53
20	2.78	1.34	−0.84	−0.54	0.17	10.10	9.76	13.24	26.13	40.77
21	2.66	1.34	−0.65	−0.75	0.27	10.45	9.76	20.21	23.00	36.59
22	1.17	1.25	0.76	−0.40	0.63	42.86	17.42	26.83	5.23	7.67

M = mean; SD = standard deviation; Sk = skewness; Ku = kurtosis.

TABLE 3 | Results of the confirmatory factor analysis.

Model	Factors	SSχ^2	df	SSχ^2/df	RMSEA [90% CI]	CFI	TLI	SRMR	WRMR
1. Ballesteros et al., 2012	1	156.361	54	2.896	0.081 [0.067, 0.096]	0.935	0.921	0.076	1.007
2. Bédard et al., 2001	2	320.371	53	6.045	0.133 [0.119, 0.147]	0.843	0.805	0.126	1.612
3. Tartaglini et al., 2020	1	356.772	119	2.998	0.084 [0.074, 0.094]	0.913	0.901	0.088	1.219
4. Martín-Carrasco et al., 2010	3	837.240	206	4.064	0.104 [0.096, 0.111]	0.806	0.782	0.118	1.683
5. Hébert et al., 2000	2	133.508	53	2.519	0.073 [0.058, 0.088]	0.962	0.953	0.064	0.893
6. Bianchi et al., 2016	3	869.989	206	4.223	0.106 [0.099, 0.113]	0.796	0.771	0.117	1.712
7. Knight et al., 2000	3	194.029	74	2.622	0.075 [0.062, 0.088]	0.930	0.914	0.084	1.063
8. Whitlatch et al., 1991	2	718.823	134	5.364	0.124 [0.115, 0.132]	0.760	0.726	0.114	1.777
9. Barreto-Osorio et al., 2015	4	649.076	203	3.197	0.088 [0.080, 0.095]	0.863	0.844	0.103	1.438
10. Rueda et al., 2017	1	176.144	65	2.710	0.077 [0.064, 0.091]	0.949	0.939	0.071	0.977
11. Knight et al., 2000	1	916.289	209	4.384	0.109 [0.102, 0.116]	0.782	0.759	0.117	1.759

SSχ^2, Chi-squared Scaling-Shifted; df, degrees of freedom; RMSEA, Root Mean Square Error of Approximation; CI, confidence interval; CFI, Comparative Fit Index; TLI, Tucker-Lewis Index; SRMR, Standardized Root Mean Square Residual; WRMR, Weighted Root Mean Square Residual.

Finally, an analysis of the ZBI was carried out from the perspective of Rasch Measurement Theory, using Andrich's Rasch model or Rating Scale. Within this analysis, the functionality of the response categories was analyzed based on the use of statistical criteria for scale optimization. In addition, the fit of the model to the data was evaluated based on three fit indices (RMSEA, TLI, and CFI), taking as acceptable values those presented in the CFA. On the other hand, the reliability of the persons and items was estimated, considering values above 0.70 adequate (Nunnally and Bernstein, 1994). Finally, the fit of the items to the model was verified using the Infit and Outfit statistics, being considered good between 0.7 and 1.3 (Wright and Linacre, 1994; Bond and Fox, 2013).

The data analysis was carried out through R version 4.1.1 (R Core Team, 2021) in the RStudio graphical user interface (RStudio Team, 2021). The tidyverse package version 1.3.1 (Wickham et al., 2019) was used for data manipulation; the Routliers package version 0.0.0.3 (Klein and Delacre, 2021) for the identification of outliers; the psych package version 2.1.9 (Revelle, 2021) for descriptive analysis; the lavaan package version 0.6-9 (Rosseel, 2012) for the CFA; the semTools package version 0.5-5 (Jorgensen et al., 2021) to estimate the reliability, AVE, and HTMT; the CTT package version 2.3.3 (Willse, 2018) to estimate the polyserial item-rest correlation; the MBESS package version 4.8.0 (Kelley, 2020) for the estimation of CIs for reliability coefficients; and the mirt package version 1.34 (Chalmers, 2012) for Rasch analysis.

RESULTS

Item Analysis

The descriptive statistics of the items (**Table 2**) indicated that item 20 presented the highest mean score (M = 2.78), while item 5 reported the lowest mean (M = 0.39). Regarding the variability of the responses, item 14 showed the highest (SD = 1.50) and item 5 the lowest (SD = 0.80). Regarding the levels of skewness and kurtosis, items 4 and 5 presented values higher than 2, indicating a slight distortion of the data concerning a normal distribution. On the other hand, most of the items presented floor and ceiling effects, where more than 15% of the responses were concentrated in response options 0 or 4. Regarding the discriminative ability of the items, all showed a polyserial item-rest correlation coefficient above 0.20, except item 20, which had a value slightly below the indicated criterion.

Validity Evidence Based on the Internal Structure

Nineteen different models were tested that differed from each other in terms of the number of factors, number of items, and the ordering of items in the factor structures. Five models did not converge because their covariance matrix of latent variables is not positive definite: Özlü et al. (2009); Ko et al. (2008), Flynn and Knight (2011); Chattat et al. (2011), and James et al. (2021). Additionally, three models were not interpreted as some estimated observed variable variances are negative: Montero et al. (2014); Lu et al. (2009), Yoon and Robinson (2005). Thus, 11 models were estimated and interpreted (**Table 3**).

According to the results presented in **Table 3**, the model of Hébert et al. (2000) and Rueda et al. (2017) has the best goodness-of-fit indices (χ^2/df < 5, CFI > 0.90, TLI > 0.90, RMSEA < 0.08, SRMR < 0.08, and WRMR < 1). The 12-item model of Hébert et al. (2000) presented two related factors. The Personal strain factor was composed of three items (items 9, 17, and 18) and the Role strain factor had nine items (items 2, 3, 6, 7, 10, 11, 12, 13, and 22), where the correlation between the factors was 0.998 and the factor loadings were between 0.350 (item 18) and 0.787 (item 12). On the other hand, the 13-item model of Rueda et al. (2017) presented a unidimensional structure (items 2, 3, 6, 9, 9, 10, 11, 12, 12, 13, 16, 17, 18, 19, and 22), with factor loadings ranging from 0.332 to 0.788 (**Table 4**).

Reliability

Reliability was analyzed through the internal consistency method with the ordinal omega coefficient. The two-factor related model of Hébert et al. (2000) presented reliability problems in the Personal strain factor (ω = 0.546), unlike the Role strain factor (ω = 0.820). Thus, the Hébert et al.'s (2000) model was discarded

TABLE 4 | Factor loading and results of Andrich's Rasch model.

Item	Factor loading	Outfit	Infit
2	0.522	1.010	1.080
3	0.722	0.661	0.703
6	0.639	0.834	1.040
9	0.699	0.855	0.964
10	0.636	0.805	0.988
11	0.665	0.852	0.875
12	0.778	0.720	0.940
13	0.788	0.732	0.839
16	0.332	1.180	1.160
17	0.725	0.793	0.865
18	0.398	1.180	1.200
19	0.509	1.080	1.030
22	0.660	0.822	0.882

TABLE 5 | Matrix of inter-item polychoric correlations.

Item	2	3	6	9	10	11	12	13	16	17	18	19	22
2	–												
3	0.54	–											
6	0.25	0.44	–										
9	0.33	0.51	0.47	–									
10	0.38	0.50	0.40	0.45	–								
11	0.38	0.43	0.40	0.55	0.33	–							
12	0.31	0.50	0.49	0.53	0.48	0.56	–						
13	0.28	0.45	0.52	0.51	0.48	0.51	0.69	–					
16	0.11	0.23	0.26	0.21	0.11	0.20	0.17	0.17	–				
17	0.33	0.42	0.49	0.52	0.39	0.53	0.55	0.62	0.37	–			
18	0.18	0.23	0.32	0.19	0.34	0.24	0.26	0.32	0.42	0.23	–		
19	0.07	0.34	0.29	0.27	0.36	0.33	0.31	0.52	0.20	0.40	0.36	–	
22	0.37	0.55	0.42	0.50	0.47	0.36	0.56	0.40	0.17	0.43	0.13	0.35	–

for further analyses. On the other hand, the unidimensional model of Rueda et al. (2017) presented a good level of reliability of the scores (ω = 0.871; 95% CI: 0.842, 0.902). Complementarily, inter-item polychoric correlations were analyzed (**Table 5**), where the coefficients were found to be between 0.07 and 0.69 (M = 0.37, SD = 0.14).

Evidence Based on Relations to Other Variables

Regarding convergent evidence, the model of the relationship between the ZBI and the CASE presented an adequate fit (χ^2 = 368.820, df = 188, χ^2/df = 1.962, CFI = 0.942, TLI = 0.935, RMSEA = 0.058 [90% CI: 0.049, 0.067], SRMR = 0.081, and WRMR = 1.068). The relationship between ZBI and CASE was significant and of strong degree (r = 0.667; $p < 0.001$; r^2 = 0.445), sharing 45% of their variability. On the other hand, the AVE of the ZBI was 0.404, considered an adequate value for a unidimensional structure and a good level of reliability of the scores (Moral, 2019).

Regarding the discriminant evidence, the HTMT ratio of the ZBI with the other instruments was 0.607, considered an acceptable level, lower than 0.85. In addition, the square root of the AVE for the ZBI (0.634) was higher than the correlation between the ZBI and CASE (0.667). Therefore, considering the results obtained, it is possible to conclude that the ZBI scores have evidence of validity based on the relationship with other variables (convergent and discriminant evidence).

Andrich Rasch Model (Rating Scale)

The unidimensional model of Rueda et al. (2017) presented a good overall fit to Andrich's Rasch model (RMSEA = 0.078, TLI = 0.930, CFI = 0.899). Likewise, the fit values for the items, Outfit, and Infit were found to be between 0.70 and 1.30 (**Table 4**), which were considered satisfactory. On the other hand, Andrich's Rasch model allowed the estimation of reliability under this perspective, presenting marginal reliability of 0.858 (Thissen and Wainer, 2001) and empirical reliability of 0.808, both considered good levels of reliability.

DISCUSSION AND CONCLUSION

The caregiver burden is a common event among people caring for others with disabilities. Zarit et al. (1980) proposed the ZBI, as an instrument to evaluate this experience. To date, different versions of this measure have been developed for some conditions (i.e., dementia, Alzheimer, and cognitive complaints). The present study was conducted to assess the psychometric properties of the ZBI in Peruvian caregivers of people with intellectual disabilities. To our knowledge, this is the first study in this sample. To this end, we evaluated the factorial structure of ZBI, its internal consistency, and its association to other variables.

Internal Structure

Regarding its internal structure, we assess the dimensionality of the previous nineteen models. Results from CFA indicated that Whitlatch et al.'s (1991) two-factor with 22 items model provided a poor fit to the data. These results are consistent with other studies (Knight et al., 2000; Li et al., 2018; Landfeldt et al., 2019). Thus, we opted to try other solutions. In the current sample, we found a significant improvement for Hébert et al.'s (2000) and Rueda et al.'s (2017) model, which have single- and two-factor structures, respectively. However, our data fit slightly better for Rueda et al.'s (2017) model.

Similar findings were obtained by Tartaglini et al. (2020), who found that the fit of the model improved by reducing to 17 items loaded on one factor, so they remained five more items (i.e., item 1: "Feel your relative asks for more help than he/she needs," item 4: "Feel embarrassed over your relative's behavior," item 5: "Feel angry when you are around your relative," item 8: "Feel your relative is dependent on you," and item 14: "Feel that your relative seems to expect you to take care of him/her as if you were the only one he/she could depend on"). Four of these five items had the weakest factor loadings in Tartaglini et al. (2020), and four of five of these items were also eliminated for Rueda et al.'s (2017) model. Regarded item content, Rasch modeling confirmed that the data fit well for Rueda et al.'s (2017) model. None of the items reported severe misfit, which demonstrated unidimensionality. In addition, the

Andrich Rasch model showed a better fit than the 12-item version of Pinyopornpanish et al. (2020), which did not demonstrate unidimensionality.

On the other hand, the results also showed a poor fit for the three and four-factor models. This finding contrasts with previous studies that revealed the ZBI is multidimensional. In this sense, Pinyopornpanish et al.'s (2020) validation study found that three-factor and four-factor provided a better fit to the data for ZBI-22 than a unidimensional model. It is noteworthy that the authors allowed the covariance of item 11 (Feel you do not have as much privacy as you) and item 12 (Feel your social life has suffered due to caring for your relative) residuals for all models (i.e., one-factor, two-factor, three-factor, and four-factor), which was needed to reach an acceptable fit. However, the specification of correlations between residuals may hide a misspecified model or a bad internal structure, showing an increase in goodness-of-fit indices, which would not contribute to the understanding of the model and the measurement of the construct (Dominguez-Lara, 2019).

Therefore, it can be concluded that, in informal caregivers of persons with intellectual disabilities, the use of a short version of the ZBI is better than the full 22-item version (ZBI-22). Specifically, in this study, the 13-item version proposed by Rueda et al. (2017) is suggested. The nine items that presented problems (1, 4, 5, 7, 8, 8, 14, 15, 20, and 21) do not contribute to the measurement of the construct "burden." Item 1 (Feel that your relative asks for more help than he/she needs) because people with intellectual disabilities require a higher level of support for certain tasks, caregivers may not consider that these people exaggerate in their demands. Items 4 (Feel embarrassed over your relative's behavior), 5 (Feel angry when you are around your relative), and 7 (Afraid what the future holds for your relative) were shown to be problematic as they link caregivers' emotions to the present or future of people with intellectual disabilities, this possibly because of the familial bonding that occurs in most of them.

The situation pointed out in the previous items may be the same as with items 20 (Feel you should be doing more for your relative) and 21 (Feel you could do a better job in caring for your relative), related to feelings of guilt and inferiority on the part of the caregivers. Items 8 (Feel your relative is dependent on you), and 14 (Feel that your relative seems to expect you to take care of him/her as if you were the only one he/she could depend on) are related to the dependence of people with intellectual disabilities and caregivers, which is assumed in most cases, because of the family bond. Finally, item 15 (Feel that you do not have enough money to take care of your relative in addition to the rest of your expenses) is the one that relates little to the rest of the ZBI items, as it refers to a socio-economic aspect of the caregiver, so it seems to be a very specific issue within the instrument.

Reliability

Regarding the reliability of the ZBI, we found that internal consistency for the personal train factor of Hébert et al.'s (2000) model was particularly low (<0.60), In contrast, the reliability for Rueda et al.'s (2017) model was adequate (>0.80). Our finding supports that, in this Peruvian sample, it is not necessary to split the ZBI into two distinct factors. These findings are very similar to the Argentinian study (Tartaglini et al., 2020) and to the Colombian sample (Rueda et al., 2017). The level of reliability obtained in the present study (ω = 0.87) was similar to that reported by Barros et al. (2019) with the 22-item version of the ZBI (α = 0.90). Likewise, the reliability estimation through the Andrich Rasch model also showed acceptable values (>0.80), similar to those obtained in the 12-item version of Pinyopornpanish et al. (2020). The results allow us to conclude that the ZBI scores present a good level of reliability from the Classical Test Theory and Rasch Measurement Theory.

Relationship With Other Variables

According to theory, people who are caregivers tend to experience various emotional and social problems (Sherwood et al., 2005). In this sample, caregiver burden displayed a significant positive association with a measure of the risk of mistreatment, a result that is consistent with previous studies (Gimeno et al., 2021), which showed a strong correlation between ZBI and CASE (>0.80). In addition, the present study found a higher correlation between the CASE and ZBI total scores (r = 0.667) compared to that reported by Rivera-Navarro et al. (2018), who showed a moderate correlation (r = 0.486) in a sample of caregivers of people with dementia. On the other hand, the AVE of the ZBI was 0.404, adequate for the context of the study (Moral, 2019), and the discriminant evidence indicators showed acceptable levels. Thus, the ZBI scores present convergent and discriminant evidence.

Strengths and Limitations

In interpreting the results from the study, several limitations are noteworthy. First, the sample was predominately female, so we were not able unable to examine the measurement invariance across gender. However, previous research has suggested that levels of experienced burden are different between women and men (Lai, 2012; Lin et al., 2017). Second, the burden was evaluated using a self-reported instrument. Thus, results could be affected by social desirability or memory biases (Althubaiti, 2016). Future research may also include other strategies as in-depth qualitative interviews. Third, we employed a convenience sample, so the results are not necessarily representative of the population. Fourth, the study design was cross-sectional, and it was not viable to assess test-retest reliability and predictive validity. Future studies using longitudinal designs could make available more useful information.

Despite the limitations, our study is the first to apply the ZBI to caregivers of people with intellectual disabilities. Moreover, the present work is the first to validate the Peruvian version of the ZBI. Furthermore, the unidimensional structure of the ZBI allows researchers and clinicians to use it and obtain an overall score easily, due to it consists of only 13 items. From a methodological perspective, this study assessed the internal structure using both Classical Test Theory and Rasch model and provided a model that explains the relationship of the ZBI with other psychological constructs using SEM. In this sense, our findings support the robust validity and reliability of the ZBI.

Conclusion

The ZBI is a commonly used tool in international caregiver burden investigation. The key contribution of this study relies on the validation of the ZBI in a sample of caregivers of people with intellectual disabilities. Analysis of internal structure validity using CFA and Rasch modeling supports a unidimensional structure. Hence, this version of the ZBI allows for a time-efficient and useful assessment of the burden.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Universidad César Vallejo. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AB-C and AT-M contributed to the conceptualization and design of the methodology of the study. AB-C provided study materials and acquired financial funding support for the project. AT-M organized the database and performed the data curation and formal analysis. AB-C and RP-A wrote the first draft of the manuscript. AB-C, RP-A, and AT-M wrote sections of the manuscript. All authors contributed to the manuscript revision, read, and approved the submitted version.

REFERENCES

Aguinis, H., Gottfredson, R. K., and Joo, H. (2013). Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organ. Res. Methods* 16, 270–301. doi: 10.1177/1094428112470848

Al-Rawashdeh, S. Y., Lennie, T. A., and Chung, M. L. (2016). Psychometrics of the Zarit Burden Interview in Caregivers of Patients with Heart Failure. *J. Cardiovasc. Nurs.* 31, E21–E28. doi: 10.1097/JCN.0000000000000348

Althubaiti, A. (2016). Information Bias in Health Research: definition, Pitfalls, and Adjustment Methods. *J. Multidiscip. Healthc.* 9, 211–217. doi: 10.2147/JMDH.S104807

American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME] (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Ato, M., López, J. J., and Benavente, A. (2013). A Classification System for Research Designs in Psychology. *An. Psicol.* 29, 1038–1059. doi: 10.6018/analesps.29.3.178511

Ballesteros, J., Santos, B., González-Fraile, E., Muñoz-Hermoso, P., Domínguez-Panchón, A. I., and Martín-Carrasco, M. (2012). "Unidimensional 12-Item Zarit Caregiver Burden Interview for the Assessment of Dementia Caregivers' Burden Obtained by Item Response Theory. *Value Health* 15, 1141–1147. doi: 10.1016/j.jval.2012.07.005

Barreto-Osorio, R. V., Campos de Aldana, M. S., Carrillo-Gonzàlez, G. M., Coral-Ibarra, R., Chaparro-Díaz, L., Duran-Parra, M., et al. (2015). Entrevista Percepción de Carga Del Cuidado de Zarit: pruebas Psicométricas Para Colombia. *Aquichan* 15, 368–380. doi: 10.5294/aqui.2015.15.3.5

Barros, A. L. O., Mancia, G., Oliveira, A., and Botti, M. T. (2019). Quality of Life and Burden of Caregivers of Children and Adolescents with Disabilities. *Spec. Care Dentist.* 39, 380–388. doi: 10.1111/scd.12400

Bédard, M., Molloy, D. W., Squire, L., Dubois, S., Lever, J. A., and O'Donnell, M. (2001). The Zarit Burden Interview: a New Short Version and Screening Version. *Gerontologist* 41, 652–657. doi: 10.1093/geront/41.5.652

Bianchi, M., Flesch, L. D., da Costa Alves, E. V., Batistoni, S. S. T., and Neri, A. L. (2016). Zarit Burden Interview Psychometric Indicators Applied in Older People Caregivers of Other Elderly. *Rev. Lat.-Am. Enferm.* 24:e2835. doi: 10.1590/1518-8345.1379.2835

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., and Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: a Primer. *Front. Public Health* 6:149. doi: 10.3389/fpubh.2018.00149

Boluarte, A. (2019). Factores Asociados a la Calidad de Vida en Personas con Discapacidad Intelectual. *Interdisciplinaria* 36, 187–202. doi: 10.16888/interd.36.1.13

Bond, T. G., and Fox, C. M. (2013). *Applying the Rasch Model: fundamental Measurement in the Human Sciences*. Hove, East Sussex: Psychology Press.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*, 2nd Edn. New York, NY: Guilford Press.

Chalmers, R. P. (2012). Mirt: a Multidimensional Item Response Theory Package for the R Environment. *J. Stat. Softw.* 48, 1–29. doi: 10.18637/jss.v048.i06

Chattat, R., Cortesi, V., Izzicupo, F., Del Re, M. L., Sgarbi, C., Fabbo, A., et al. (2011). The Italian Version of the Zarit Burden Interview: a Validation Study. *Int. Psychogeriatr.* 23, 797–805. doi: 10.1017/S1041610210002218

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, S. R., Holloway, S. D., Domínguez-Pareto, I., and Kuppermann, M. (2014). Receiving or Believing in Family Support? Contributors to the Life Quality of Latino and Non-Latino Families of Children with Intellectual Disability. *J. Intellect. Disabil. Res.* 58, 333–345. doi: 10.1111/jir.12016

Conti, A., Clari, M., Garrino, L., Maitan, P., Scivoletto, G., Cavallaro, L., et al. (2019). Adaptation and Validation of the Caregiver Burden Inventory in Spinal Cord Injuries (CBI-SCI). *Spinal Cord* 57, 75–82. doi: 10.1038/s41393-018-0179-7

Crespo, M., and Rivas, M. T. (2015). La Evaluación de La Carga Del Cuidador: una Revisión Más Allá de La Escala de Zarit. *Clín. Salud* 26, 9–15. doi: 10.1016/j.clysa.2014.07.002

Cuenot, M. (2018). Clasificación Internacional del Funcionamiento, de la Discapacidad y de la Salud. *EMC - Kinesiterapia - Medicina Física* 39, 1–6. doi: 10.1016/S1293-2965(18)88602-9

da Cruz, C. A. (2010). Adaptação e Validação Da Escala de Sobrecarga Do Cuidador de Zarit. *Ref. Rev. Enferm.* 2, 9–16.

DiStefano, C., Liu, J., Jiang, N., and Shi, D. (2018). Examination of the Weighted Root Mean Square Residual: evidence for Trustworthiness? *Struct. Equ. Modeling* 25, 453–466. doi: 10.1080/10705511.2017.1390394

Dominguez-Lara, S. (2019). Correlación entre Residuales en Análisis Factorial Confirmatorio: una Breve Guía para su Uso e Interpretación. *Interacciones* 5:e207. doi: 10.24016/2019.v5n3.207

Duarte, J. M., and García-Horta, J. B. (2016). Igualdad, Equidad de Género y Feminismo, una Mirada Histórica a la Conquista de los Derechos de las Mujeres. *Rev. CS* 18, 107–158. doi: 10.18046/recs.i18.1960

Espín, A. M. (2008). Caracterización Psicosocial de Cuidadores Informales de Adultos Mayores con Demencia. *Rev. Cub. Salud Pública* 34, 1–12. doi: 10.1590/s0864-34662008000300002

Farmer, C., Thienemann, M., Leibold, C., Kamalani, G., Sauls, B., and Frankovich, J. (2018). Psychometric Evaluation of the Caregiver Burden Inventory in Children and Adolescents with PANS. *J. Pediatr. Psychol.* 43, 749–757. doi: 10.1093/jpepsy/jsy014

Flora, D. B. (2020). Your Coefficient Alpha Is Probably Wrong, but Which Coefficient Omega Is Right? A Tutorial on Using R to Obtain Better Reliability Estimates. *Adv. Methods Pract. Psychol. Sci.* 3, 484–501. doi: 10.1177/2515245920951747

Flynn, C. V., and Knight, B. G. (2011). Confirmatory Factor Analysis of a Brief Version of the Zarit Burden Interview in Black and White Dementia Caregivers. *Gerontologist* 51, 453–462. doi: 10.1093/geront/gnr011

Fornell, C., and Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *J. Mark. Res.* 18, 39–50. doi: 10.2307/3151312

Gimeno, I., Val, S., and Cardoso, M. J. (2021). "Relation among Caregivers' Burden, Abuse and Behavioural Disorder in People with Dementia. *Int. J. Environ. Res. Public Health* 18:1263. doi: 10.3390/ijerph18031263

Hébert, R., Bravo, G., and Préville, M. (2000). Reliability, Validity and Reference Values of the Zarit Burden Interview for Assessing Informal Caregivers of Community-Dwelling Older Persons with Dementia. *Can. J. Aging* 19, 494–507. doi: 10.1017/S0714980800012484

Henseler, J., Ringle, C. M., and Sarstedt, M. (2015). A New Criterion for Assessing Discriminant Validity in Variance-Based Structural Equation Modeling. *J. Acad. Mark. Sci.* 43, 115–135. doi: 10.1007/s11747-014-0403-8

Instituto Nacional de Estadística e Informática [INEI] (2019). *Perfil Sociodemográfico de la Población con Discapacidad, 2017*. Lima: Instituto Nacional de Estadística e Informática.

James, K., Chin-Bailey, C., Holder-Nevins, D., Thompson, C., Donaldson-Davis, K., and Eldemire-Shearer, D. (2021). Zarit Burden Interview among Caregivers of Community-Dwelling Older Adults in a Caribbean Setting (Jamaica): reliability and Factor Structure. *Health Soc. Care Community* 29, e79–e88. doi: 10.1111/hsc.13244

Jarrett, W. H. (1985). Caregiving within Kinship Systems: is Affection Really Necessary? *Gerontologist* 25, 5–10. doi: 10.1093/geront/25.1.5

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., and Rosseel, Y. (2021). *SemTools: useful Tools for Structural Equation Modeling*. Available Online at: https://cran.r-project.org/package=semTools (accessed September 21, 2021).

Keith, T. Z. (2019). *Multiple Regression and beyond: an Introduction to Multiple Regression and Structural Equation Modeling*, 3rd Edn. New York, NY: Routledge, doi: 10.4324/9781315162348

Kelley, K. (2020). *MBESS: the MBESS R Package*. Available Online at: https://cran.r-project.org/package=MBESS (accessed September 21, 2021).

Kerlinger, F. N., and Lee, H. B. (2000). *Foundations of Behavioral Research*, 4th Edn. Fort Worth, TX: Harcourt College Publishers.

Khan, A., Adil, A., Ameer, S., and Shujja, S. (2020). Caregiver Abuse Screen for Older Adults: urdu Translation, Validation, Factorial Equivalence, and Measurement Invariance. *Curr. Psychol.* 1–11. doi: 10.1007/s12144-020-00894-y [Epub ahead of print].

Kim, J., Kim, H., Park, S., Yoo, J., and Gelegjamts, D. (2021). Mediating Effects of Family Functioning on the Relationship between Care Burden and Family Quality of Life of Caregivers of Children with Intellectual Disabilities in Mongolia. *J. Appl. Res. Intellect. Disabil.* 34, 507–515. doi: 10.1111/jar.12814

Klein, O., and Delacre, M. (2021). *"Routliers: robust Outliers Detection*. Available Online at: https://cran.r-project.org/package=Routliers (accessed September 21, 2021).

Knight, B. G., Fox, L. S., and Chou, C.-P. (2000). Factor Structure of the Burden Interview. *J. Clin. Geropsychol.* 6, 249–258. doi: 10.1023/A:1009530711710

Ko, K.-T., Yip, P.-K., Liu, S.-I., and Huang, C.-R. (2008). Chinese Version of the Zarit Caregiver Burden Interview: a Validation Study. *Am. J. Geriatr. Psychiatry* 16, 513–518. doi: 10.1097/JGP.0b013e318167ae5b

Lai, D. W. L. (2012). Effect of Financial Costs on Caregiving Burden of Family Caregivers of Older Adults. *SAGE Open* 2, 1–14. doi: 10.1177/2158244012470467

Landfeldt, E., Mayhew, A., Straub, V., Bushby, K., Lochmüller, H., and Lindgren, P. (2019). Psychometric Properties of the Zarit Caregiver Burden Interview Administered to Caregivers to Patients with Duchenne Muscular Dystrophy: a Rasch Analysis. *Disabil. Rehabil.* 41, 966–973. doi: 10.1080/09638288.2017.1416501

Leys, C., Delacre, M., Mora, Y. L., Lakens, D., and Ley, C. (2019). How to Classify, Detect, and Manage Univariate and Multivariate Outliers, with Emphasis on Pre-Registration. *Int. Rev. Soc. Psychol.* 32:5. doi: 10.5334/irsp.289

Leys, C., Klein, O., Dominicy, Y., and Ley, C. (2018). Detecting Multivariate Outliers: use a Robust Variant of the Mahalanobis Distance. *J. Exp. Soc. Psychol.* 74, 150–156. doi: 10.1016/j.jesp.2017.09.011

Li, R., Chong, M. S., Chan, P. C. M., Tay, B. G. L., Ali, N. B., and Lim, W. S. (2018). Worry about Caregiving Performance: a Confirmatory Factor Analysis. *Front. Med.* 5:79. doi: 10.3389/fmed.2018.00079

Lima-Rodríguez, J. S., Baena-Ariza, M. T., Domínguez-Sánchez, I., and Lima-Serrano, M. (2018). Discapacidad Intelectual en Niños y Adolescentes: influencia en la Familia y la Salud Familiar. Revisión Sistemática. *Enferm. Clín.* 28, 89–102. doi: 10.1016/j.enfcli.2017.10.005

Lin, C.-Y., Ku, L.-J. E., and Pakpour, A. H. (2017). Measurement Invariance across Educational Levels and Gender in 12-Item Zarit Burden Interview (ZBI) on Caregivers of People with Dementia. *Int. Psychogeriatr.* 29, 1841–1848. doi: 10.1017/S1041610217001417

Lu, L., Wang, L., Yang, X., and Feng, Q. (2009). Zarit Caregiver Burden Interview: development, Reliability and Validity of the Chinese Version. *Psychiatry Clin. Neurosci.* 63, 730–734. doi: 10.1111/j.1440-1819.2009.02019.x

Martín-Carrasco, M., Otermin, P., érez-Camo, V. P., Pujol, J., Agüera, L., Martín, M. J., et al. (2010). EDUCA Study: psychometric Properties of the Spanish Version of the Zarit Caregiver Burden Scale. *Aging Ment. Health* 14, 705–711. doi: 10.1080/13607860903586094

Martín-Carrasco, M., Salvadó, I., Nadal, S., Miji, L., Rico, J. M., and Lanz, P. (1996). Adaptación para Nuestro Medio de la Escala de Sobrecarga del Cuidador (Caregiver Burden Interview) de Zarit. *Rev. Gerontol.* 6, 338–346.

McHorney, C. A., and Tarlov, A. R. (1995). Individual-Patient Monitoring in Clinical Practice: are Available Health Status Surveys Adequate? *Qual. Life Res.* 4, 293–307. doi: 10.1007/BF01593882

Mikulincer, M., and Shaver, P. R. (2005). Attachment Theory and Emotions in Close Relationships: exploring the Attachment-Related Dynamics of Emotional Reactions to Relational Events. *Pers. Relationsh.* 12, 149–168. doi: 10.1111/j.1350-4126.2005.00108.x

Monreal, A., and Prieto, G. (2017). Análisis del Test de Detección del Cuidador Quemado con el Modelo de Rasch. *Escr. Psicol.* 10, 116–125. doi: 10.5231/psy.writ.2017.1904

Montero, X., Jurado, S., Valencia, A., Méndez, J., and Mora, I. (2014). Escala de Carga del Cuidador de Zarit: evidencia de Validez en México. *Psicooncología* 11, 71–85. doi: 10.5209/rev_PSIC.2014.v11.n1.44918

Montorio, I., Fernández, M. I., López, A., and Sánchez, M. (1998). La Entrevista de Carga del Cuidador. Utilidad y Validez del Concepto de Carga. *An. Psicol.* 14, 229–248.

Moral, J. (2019). Revisión de los Criterios Para Validez Convergente Estimada a través de la Varianza Media Extraída. *Psychologia* 13, 25–41. doi: 10.21500/19002386.4119

Mundfrom, D. J., Shaw, D. G., and Ke, T. L. (2005). Minimum Sample Size Recommendations for Conducting Factor Analyses. *Int. J. Test.* 5, 159–168. doi: 10.1207/s15327574ijt0502_4

Novak, M., and Guest, C. (1989). Application of a Multidimensional Caregiver Burden Inventory. *Gerontologist* 29, 798–803. doi: 10.1093/geront/29.6.798

Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*, 3rd Edn. New York, NY: McGraw-Hill.

Orfila, F., Coma-Solé, M., Cabanas, M., Cegri-Lombardo, F., Moleras-Serra, A., and Pujol-Ribera, E. (2018). Family Caregiver Mistreatment of the Elderly: prevalence of Risk and Associated Factors. *BMC Public Health* 18:167. doi: 10.1186/s12889-018-5067-8

O'Rourke, N., and Tuokko, H. A. (2003). Psychometric Properties of an Abridged Version of the Zarit Burden Interview within a Representative Canadian Caregiver Sample. *Gerontologist* 43, 121–127. doi: 10.1093/geront/43.1.121

Ortiz-Rubio, A., Torres-Sánchez, I., Cabrera-Martos, I., Rodríguez-Torres, J., López-López, L., Prados-Román, E., et al. (2021). The Caregiver Burden Inventory as a Sleep Disturbance Screening Tool for Parents of Children with Autism Spectrum Disorder. *J. Pediatr. Nurs.* 61, 166–172. doi: 10.1016/j.pedn.2021.05.013

Özcan, N. K., Boyacioğlu, N. E., and Sertçelik, E. (2017). Reciprocal Abuse: elder Neglect and Abuse by Primary Caregivers and Caregiver Burden and Abuse in Turkey. *Arch. Psychiatr. Nurs.* 31, 177–182. doi: 10.1016/j.apnu.2016.09.011

Özlü, A., Yildiz, M., and Aker, T. (2009). A Reliability and Validity Study on the Zarit Caregiver Burden Scale. *Arch. Neuropsychiatry* 46, 38–42.

Pérez-Rojo, G., Nuevo, R., Sancho, M., and Penhale, B. (2015). Validity and Reliability of the Spanish Version of Caregiver Abuse Screen (CASE). *Res. Aging* 37, 63–81. doi: 10.1177/0164027514522275

Pinyopornpanish, K., Pinyopornpanish, M., Wongpakaran, N., Wongpakaran, T., Soontornpun, A., and Kuntawong, P. (2020). Investigating Psychometric Properties of the Thai Version of the Zarit Burden Interview Using Rasch Model and Confirmatory Factor Analysis. *BMC Res. Notes* 13:120. doi: 10.1186/s13104-020-04967-w

R Core Team (2021). *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rajabi-Mashhadi, M. T., Mashhadinejad, H., Ebrahimzadeh, M. H., Golhasani-Keshtan, F., Ebrahimi, H., and Zarei, Z. (2015). The Zarit Caregiver Burden Interview Short Form (ZBI-12) in Spouses of Veterans with Chronic Spinal Cord Injury, Validity and Reliability of the Persian Version. *Arch. Bone Jt. Surg.* 3, 56–63.

Reichenheim, M. E., Paixão, C. M. Jr., and Moraes, C. L. (2009). Reassessing the Construct Validity of a Brazilian Version of the Instrument Caregiver Abuse Screen (CASE) Used to Identify Risk of Domestic Violence against the Elderly. *J. Epidemiol. Community Health* 63, 878–883. doi: 10.1136/jech.2008.084095

Reis, M., and Nahmiash, D. (1995). Validation of the Caregiver Abuse Screen (CASE). *Can. J. Aging* 14, 45–60. doi: 10.1017/S0714980800005584

Revelle, W. (2021). *Psych: procedures for Psychological, Psychometric, and Personality Research*. Evanston, IL: Northwestern University.

Rivera-Navarro, J., Sepúlveda, R., Contador, I., Fernández-Calvo, B., Ramos, F., Tola-Arribas, M. A., et al. (2018). Detection of Maltreatment of People with Dementia in Spain: usefulness of the Caregiver Abuse Screen (CASE). *Eur. J. Ageing* 15, 87–99. doi: 10.1007/s10433-017-0427-2

Rodríguez, D. C., Bravo, K. T., and González, E. C. (2019). Percepciones de la Sobrecarga Subjetiva en Cuidadoras Principales de Niños o Adolescentes con Discapacidad Física, Cognitiva o Sensorial a través de su Experiencia en la Labor de Cuidar. *Hal* [Preprint]. Available Online at: https://hal.archives-ouvertes.fr/hal-02411611 (accessed September 21, 2021).

Rosseel, Y. (2012). Lavaan: an R Package for Structural Equation Modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

RStudio Team (2021). *RStudio: integrated Development Environment for R*. Boston, MA: RStudio, PBC.

Rueda, L. J., Ramos, J. X., and Márquez, M. (2017). Análisis de Rasch de la Escala Burden Interview de Zarit Aplicada a Cuidadores Familiares en Bucaramanga, Colombia. *Arch. Med.* 17, 17–26. doi: 10.30554/archmed.17.1.1804.2017

Ruiz, A. E., and Nava, M. G. (2012). Cuidadores: responsabilidades-Obligaciones. *Rev. Enferm. Neurol.* 11, 163–169. doi: 10.37976/enfermeria.v11i3.149

Saravia, V. T., Pinheiro, N. C., Silva, I., Athie, S., and Ramos, E. (2019). Prevalence and Factors Associated with Caregiver Abuse of Elderly Dependents: the Hidden Face of Family Violence. *Cienc. Saude Colet.* 24, 87–96. doi: 10.1590/1413-81232018241.34872016

Schalock, R. L., Luckasson, R., and Tassé, M. J. (2021). *Intellectual Disability: definition, Diagnosis, Classification, and Systems of Supports*, 12th Edn. Washington, DC: American Association on Intellectual and Developmental Disabilities.

Schmeiser, C. B., and Welch, C. J. (2006). "Test Development," in *Educational Measurement*, 4th Edn, ed. R. L. Brennan (Westport, CT: Praeger Publishers), 307–353.

Schreiner, A. S., Morimoto, T., Arai, Y., and Zarit, S. (2006). Assessing Family Caregiver's Mental Health Using a Statistically Derived Cut-off Score for the Zarit Burden Interview. *Aging Ment. Health* 10, 107–111. doi: 10.1080/13607860500312142

Schumacker, R. E., and Lomax, R. G. (2016). *A Beginner's Guide to Structural Equation Modeling*, 4th Edn. New York, NY: Routledge.

Seguí, J. D., Ortiz-Tallo, M., and De Diego, Y. (2008). Factores Asociados al Estrés del Cuidador Primario de Niños con Autismo: sobrecarga, Psicopatología y Estado de Salud. *An. Psicol.* 24, 100–105.

Seng, B. K., Luo, N., Ng, W. Y., Lim, J., Chionh, H. L., Goh, J., et al. (2010). Validity and Reliability of the Zarit Burden Interview in Assessing Caregiving Burden. *Ann. Acad. Med. Singap.* 39, 758–763.

Sherwood, P. R., Given, C. W., Given, B. A., and von Eye, A. (2005). Caregiver Burden and Depressive Symptoms: analysis of Common Outcomes in Caregivers of Elderly Patients. *J. Aging Health* 17, 125–147. doi: 10.1177/0898264304274179

Siegert, R. J., Jackson, D., Tennant, A., and Turner-Stokes, L. (2010). Factor Analysis and Rasch Analysis of the Zarit Burden Interview for Acquired Brain Injury Carer Research. *J. Rehabil. Med.* 42, 302–309. doi: 10.2340/16501977-0511

Tabachnick, B. G., and Fidell, L. S. (2019). *Using Multivariate Statistics*, 7th Edn. New York, NY: Pearson.

Tang, J. Y.-M., Ho, A. H.-Y., Luo, H., Wong, G. H.-Y., Lau, B. H.-P., Lum, T. Y.-S., et al. (2016). Validating a Cantonese Short Version of the Zarit Burden Interview (CZBI-Short) for Dementia Caregivers. *Aging Ment. Health* 20, 996–1001. doi: 10.1080/13607863.2015.1047323

Tartaglini, M. F., Feldberg, C., Hermida, P. D., Heisecke, S. L., Dillon, C., Ofman, S. D., et al. (2020). Escala de Sobrecarga del Cuidador de Zarit: análisis de sus Propiedades Psicométricas en Cuidadores Familiares Residentes en Buenos Aires, Argentina. *Neurol. Argent.* 12, 27–35. doi: 10.1016/j.neuarg.2019.11.003

Tassé, M. J., Luckasson, R., and Schalock, R. L. (2016). The Relation between Intellectual Functioning and Adaptive Behavior in the Diagnosis of Intellectual Disability. *Intellect. Dev. Disabil.* 54, 381–390. doi: 10.1352/1934-9556-54.6.381

Taub, A., Andreoli, S. B., and Bertolucci, P. H. (2004). Dementia Caregiver Burden: reliability of the Brazilian Version of the Zarit Caregiver Burden Interview. *Cad. Saúde Pública* 20, 372–376. doi: 10.1590/S0102-311X2004000200004

Thabane, L., Mbuagbaw, L., Zhang, S., Samaan, Z., Marcucci, M., Ye, C., et al. (2013). A Tutorial on Sensitivity Analyses in Clinical Trials: the What, Why, When and How. *BMC Med. Res. Methodol.* 13:92. doi: 10.1186/1471-2288-13-92

Thissen, D., and Wainer, H. (2001). *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum Associates.

Uzun, A., Sarı, S. A., and Mercan, C. (2020). Sociodemographic Characteristics, Risk Factors, and Prevalence of Comorbidity among Children and Adolescents with Intellectual Disability: a Cross-Sectional Study. *J. Ment. Health Res. Intellect. Disabil.* 13, 66–85. doi: 10.1080/19315864.2020.1727590

Vázquez, F. L., Otero, P., Simón, M. A., Bueno, A. M., and Blanco, V. (2019). Psychometric Properties of the Spanish Version of the Caregiver Burden Inventory. *Int. J. Environ. Res. Public Health* 16:217. doi: 10.3390/ijerph16020217

Ventura-León, J., and Peña-Calero, B. N. (2021). The World Should Not Revolve around Cronbach's Alpha = .70. *Adicciones* 33, 369–372. doi: 10.20882/adicciones.1576

Viladrich, C., Angulo-Brunet, A., and Doval, E. (2017). A Journey around Alpha and Omega to Estimate Internal Consistency Reliability. *An. Psicol.* 33, 755–782. doi: 10.6018/analesps.33.3.268401

Vitaliano, P. P., Russo, J., Young, H. M., Becker, J., and Maiuro, R. D. (1991). The Screen for Caregiver Burden. *Gerontologist* 31, 76–83. doi: 10.1093/geront/31.1.76

Whitlatch, C. J., Zarit, S. H., and von Eye, A. (1991). Efficacy of Interventions with Caregivers: a Reanalysis. *Gerontologist*. 31, 9–14. doi: 10.1093/geront/31.1.9

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686

Willse, J. T. (2018). *CTT: classical Test Theory Functions*. Available Online at: https://cran.r-project.org/package=CTT (accessed September 21, 2021).

World Health Organization [WHO] (2007). *International Classification of Functioning, Disability and Health: children and Youth Version (ICF-CY)*. Geneva: World Health Organization.

World Health Organization [WHO], and World Bank (2011). *World Report on Disability*. Geneva: World Health Organization.

Wright, B. D., and Linacre, J. M. (1994). Reasonable Mean-Square Fit Values. *Rasch Meas. Trans.* 8, 370–371.

Yoon, E., and Robinson, M. (2005). Psychometric Properties of the Korean Version of the Zarit Burden Interview (K-ZBI): preliminary Analyses. *J. Soc. Work Res. Eval.* 6, 75–86.

Yu, Y., Liu, Z.-W., Zhou, W., Chen, X.-C., Zhang, X.-Y., Hu, M., et al. (2018). Assessment of Burden among Family Caregivers of Schizophrenia: psychometric Testing for Short-Form Zarit Burden Interviews. *Front. Psychol.* 9:2539. doi: 10.3389/fpsyg.2018.02539

Zarit, S. H., Orr, N. K., and Zarit, J. M. (1985). *The Hidden Victims of Alzheimer's Disease: families under Stress*. New York, NY: New York University Press.

Zarit, S. H., Reever, K. E., and Bach-Peterson, J. (1980). Relatives of the Impaired Elderly: correlates of Feelings of Burden. *Gerontologist* 20, 649–655. doi: 10.1093/geront/20.6.649

Zarit, S. H., Todd, P. A., and Zarit, J. M. (1986). Subjective Burden of Husbands and Wives as Caregivers: a Longitudinal Study. *Gerontologist* 26, 260–266. doi: 10.1093/geront/26.3.260

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 22 April 2022
doi: 10.3389/fpsyg.2022.773447

Check for updates

Separate but Related: Dimensions of Healthcare Provider Social Support in Day-Treatment Oncology Units

Manuela Tomai[1*†] and Marco Lauriola[2*†]

[1]Department of Dynamic and Clinical Psychology, and Health, Sapienza University of Rome, Rome, Italy, [2]Department of Social and Developmental Psychology, Sapienza University of Rome, Rome, Italy

Social support by healthcare providers has been increasingly investigated during the past decade, but studies have made different choices concerning its measurement. To evaluate how social support from a healthcare provider impacts the perceived quality of care and patient outcomes, reliable and valid instruments capable of measuring specific aspects of the construct are needed. In study 1, we tested the factor structure and the psychometric properties of a new Healthcare Provider Social Support measure (HPSS) for oncology settings. One-hundred-sixty-two patients (89 females; *M* age = 58.97, *SD* age = 13.28) from religious and government-operated hospitals completed the HPSS during day treatment. We modeled the HPSS factor structure to represent four related aspects: Emotional, Informational, Appraisal, and Instrumental social support. Study 2 preliminarily assessed the concurrent validity of the HPSS with patient perceptions of the patient-doctor relationship. Sixty-nine patients (40 females; *M* age = 53.67, *SD* age = 13.74) completed the HPPS with scales assessing perceived doctor-patient communication and patient trust in the healthcare provider. Study 1, using Exploratory Structural Equation Modeling, showed that a bifactor model had an excellent fit. The analysis supported the use of subscale scores, which were more tenable than a single total score in terms of bifactor model indices. This conclusion was also supported by greater scalability of the subscales in a Mokken Scale Analysis. Oncology patients treated in the religious hospital perceived greater Emotional, Informational, and Instrumental social support from their healthcare provider than those treated in government-operated. Study 2 showed that patient ratings of healthcare provider social support, except Instrumental, were positively correlated with better doctor communication skills and greater trust in the physician. Multiple regression analyses showed that Informational and Emotional support provided a unique contribution to building trust in the physician, controlling for the doctor's communication skills. The study results showed that the four social support ratings were reliable and valid, sharpening the distinction between functional components in the formal healthcare system.

Keywords: social support, healthcare, day treatment, cancer patients, scale construction and validation

Edited by:
Paola Gremigni,
University of Bologna, Italy

Reviewed by:
Ilaria Setti,
University of Pavia, Italy
Cesar Merino-Soto,
Universidad de San Martin de Porres, Peru
Valeria Sebri,
European Institute of Oncology (IEO), Italy

****Correspondence:***
Manuela Tomai
manuela.tomai@uniroma1.it
Marco Lauriola
marco.lauriola@uniroma1.it

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *09 September 2021*
Accepted: *08 March 2022*
Published: *22 April 2022*

Citation:
Tomai M and Lauriola M (2022) Separate but Related: Dimensions of Healthcare Provider Social Support in Day-Treatment Oncology Units. Front. Psychol. 13:773447. doi: 10.3389/fpsyg.2022.773447

INTRODUCTION

The quality of close relationships and the sense of social connectedness are reliable predictors of health and longevity (Holt-Lunstad et al., 2017; Schetter, 2017). Not surprisingly, the WHO enlisted social support networks among the most critical determinants of health (Commission on Social Determinants of Health, 2008). Social support networks perform four main functions: Emotional, Informational, Appraisal, and Instrumental. Briefly, *Emotional* support (also called affective or attachment) refers to demonstrations of love and caring, encouragement, and empathy from which one derives a sense of security (Thoits, 2011). *Informational* support is defined as providing facts or advice that may help a person solve problems (Wang et al., 2014). *Instrumental* support is intended as a form of practical, tangible help, provided through material assistance or practical tasks (Langford et al., 1997; Wang et al., 2014). Last, *Appraisal* support consists of expressions that affirm the appropriateness of acts or statements made by another (Langford et al., 1997).

Both informal and formal networks can support people affected by chronic diseases. Thus, this article refers to social support activated or required to cope with adverse life events and chronic health problems, defined as problem-oriented social support (e.g., Suurmeijer et al., 1995). Research has consistently demonstrated several benefits that medical patients receive from their informal networks (e.g., family, friends, and relatives). In conditions like diabetes, these benefits include improving emotion regulation, coping, glycemic control, and quality of life (Van Dam et al., 2005; Strom and Egede, 2012; Hill-Briggs et al., 2021). Social support also enhanced functional status and quality of life in people with heart failure, influencing treatment and health management behaviors (Graven and Grant, 2013, 2014; Bucholz et al., 2014).

In oncology patients, social support was found to address psychological problems resulting from a poor adjustment to cancer and its treatment (Rizalar et al., 2014). For example, by fostering acceptance, positive reframing of the situation, and maintaining the patient's sense of humor, social support networks help patients be more determined in their fight against the disease and counteract helplessness-hopelessness and anxious preoccupations (Kawa, 2017; Lauriola and Tomai, 2019; Tomai et al., 2019). Moreover, several authors (Pinquart and Duberstein, 2010; Ikeda et al., 2013; Yağmur and Duman, 2016) reported significant correlations of social support with "hard" health outcomes, such as tumor initiation, cell proliferation, and life extension.

It is worth noting that social support may unintentionally hurt the patient's wellbeing. Previous research has indicated that, contrary to caregivers' intentions, patients who report unmet needs may perceive some forms of support as ineffective or troublesome (Breuer et al., 2017; Sebri et al., 2021). In particular, social support turns out unsupportive when the helper does not understand or effectively correspond to the recipient's desires (Nouman and Zanbar, 2020). For example, this can happen within close relationships, when help is not sought or when support becomes controlling, oppressive, or, conversely, too superficial and neglecting the consequences of illness (Mazzoni and Cicognani, 2016; Mazzoni et al., 2017). Therefore, paying attention to the type of support a patient needs and how that support is provided will be the best way to address patients' specific needs.

In addition to family and friends, healthcare professionals (e.g., doctors and nurses) can provide social support to patients suffering from chronic diseases, thus becoming a *formal social support network*. For instance, a supportive healthcare provider helped diabetic patients to defuse health distress and improve glycemia (Venkatesh and Weatherspoon, 2013; Wardian and Sun, 2014). Similarly, according to a recent study (Ban et al., 2021), the medical staff reduced cancer patients' fear of illness progression by providing social support. Indeed, identifying patients at risk for psychosocial vulnerability due to low healthcare support appears to be a relevant health outcome (Usta, 2012).

How do formal social support networks work in oncology units? Are there similar functions that informal and formal networks perform for cancer patients? Healthcare social support functions have been less extensively studied than informal network ones. Nevertheless, given the increasing interest in improving healthcare quality and patient experience, constructs similar to social support functions have been used in medical settings (e.g., Donabedian, 2005; Beattie et al., 2015; Hancock et al., 2020). For instance, healthcare quality indicators include ratings of physicians' interpersonal skills and information provision (Brédart et al., 2005a,b). Trust in oncologists reflects the patients' beliefs about the healthcare provider's ability to provide appropriate, reliable, and hopefully successful treatment (Müller et al., 2014). Communication is also considered a core clinical skill for establishing an excellent patient–doctor relationship (Malley and Fernández, 2010). When clinicians are better communicators, they can effectively convey information to patients, encouragement, and tangible support (Street et al., 2009). In sum, a health professional's interpersonal competence and patient trust in the healthcare provider might be associated with one's perception of social support. Still, they are not part of the construct as defined in psychosocial research (Barrera and Ainlay, 1983).

Where scholars have explicitly addressed healthcare social support, methodological approaches to obtaining reliable and valid measures have been varied and scattered. Previous research lacked a clear connection with the fourfold structure of the construct. For instance, some studies assessed multiple aspects of social support but used a single global score (Katz et al., 2003; Reynolds and Perrin, 2004). This choice, however, does not make fine-grained distinctions between how the healthcare provider can support the patient and improve the therapeutic relationship. Other studies focused either on emotional support (Kuuppelomäki, 2003; Wenrich et al., 2003; Ansmann et al., 2012) or informational support (Rutten et al., 2005). Underlying this approach, emotional and informational supports are viewed as functionally independent and not interchangeable. Another original approach combined two questions for each type of social support function into a single dichotomous indicator (Arora and Gustafson, 2009), assuming that healthcare provider social support could be a categorical variable. Unclear boundaries exist between informal networks' social support functions and other proxy constructs used in medical research (e.g., communication skills).

Healthcare social support depends not only on individual actors (e.g., doctors or nurses) but also on the hospital's mission and organizational culture. Ansmann et al. (2014), for example, emphasized the importance of hospital characteristics in fostering a better patient-doctor relationship. Hospitals in which the local culture emphasized cost control typically reported greater dissatisfaction with physicians' interpersonal skills and information provision (Zhou et al., 2011). By contrast, non-profit hospitals imbibed in a religious culture promote a more inclusive and respectful atmosphere (Whitley, 2012). Religious hospitals have traditionally been reputed to provide a higher quality of care than government-operated hospitals (Fleming, 1981). Indeed, the organization's charitable mission to serve and care for the person, along with a religious institutional identity, promotes the humanization of health care (e.g., Reinikka and Svensson, 2010; Calegari et al., 2015). Thus, religious hospitals can offer close and sensitive listening to the patient's narratives and more compassionate care services, including care for the poor and vulnerable (White and Begun, 1998; White et al., 2010). Users themselves describe religious hospitals as more reliable and attractive than government-operated ones (Seemann et al., 2015). Considering this literature, a healthcare provider social support measure could also be a valuable tool to compare hospitals, outpatient clinics, particular services, or departments.

The Present Study

The role of social support from healthcare providers is a challenging and understudied area of research. There is a need to understand how and when the support provided by health professionals may influence the quality of care and which social support function has the most significant impact on the quality-of-care cancer patients receive in oncology centers. The present study sought to develop a *Healthcare Provider Social Support* scale (HPSS), sharpening the distinction between Emotional, Informational, Instrumental, and Appraisal functions in the formal oncology care system. These domains were inspired by Cutrona and Russell (1987). However, we did not develop items related to "Opportunity for Nurturance" (i.e., the sense of being needed by others for their wellbeing) and "Social Integration" (i.e., the importance of belonging to a group that shares similar interests). These forms of social support primarily belong to informal networks (Cutrona and Russell, 1987).

To the best of our knowledge, no previous study has developed a psychometric tool that reflects the construct's fourfold model, which proved to be valid for assessing patients' social support from their informal network. Study 1 was designed to preliminary assess the HPSS dimensionality and reliability in a sample of patients recruited from 2 day-treatment oncology units. According to the multidimensionality of the construct, we hypothesized that at least four factors are identified in the factor analysis of a reliable item set. These factors should correspond to perceived Emotional, Informational, Instrumental, and Appraisal support. A subsequent study used an independent patient sample to assess the relationships between HPSS ratings and other related constructs in the nomological network, such as doctor communication skills and the patient's trust in the physician. We hypothesized that greater healthcare provider support perceptions are associated with patient-centered communication and greater trust in the physician.

STUDY 1: FACTOR STRUCTURE AND SCALABILITY OF THE HPSS

Investigating the factor structure of a new scale is vital to establishing its psychometric properties. Confirmatory Factor Analysis (CFA) is commonly used for this purpose. Nevertheless, CFA has recently complained about failing to account for the imperfect nature of items, as evidenced by significant item correlations with non-target constructs or cross-loadings on more than a factor (Marsh et al., 2014; Morin et al., 2016). As a result, CFA may produce inaccurate estimates of factor correlations, necessitating general or method factors to achieve an acceptable fit (Morin et al., 2016; Joshanloo et al., 2017). In the present study, we used Exploratory Structural Equation Modeling (ESEM), a new approach that overcomes the limitations of CFA (Asparouhov and Muthén, 2009). In ESEM, cross-loadings on non-target factors are allowed in addition to primary loadings on target factors. Thus, ESEM is more efficient than CFA in the estimation of complex models because it provides more accurate estimates of factor loadings and correlations, preventing artificial inflation of factor loadings on the general factor (Marsh et al., 2014; Morin et al., 2016). We further examined the HPSS dimensionality and reliability using Mokken Scaling Analysis (MSA) to provide a sensitivity analysis for ESEM results. As shown in **Figure 1**, we tested the following factor models, each based on different theoretical assumptions. Like previous studies (Katz et al., 2003; Reynolds and Perrin, 2004), the unifactorial model assumes that all HPSS items measure the generalized patient's perception of the healthcare provider's support (**Figure 1A**). The four-factor model is consistent with the view that patients could discriminate between ways a doctor had supported them (**Figure 1B**). Assuming correlated factors would imply that the four social support functions were somewhat interrelated. Small factor correlations would show that different types of support are distinct but related aspects. Moderate factor correlations would indicate substantial overlap, suggesting a hierarchical arrangement of factors. In a bifactor model, every item is targeted to load primarily on general and specific factors (**Figure 1C**). For example, an item describing an emotionally supportive act is hypothesized to load on a generalized support perception factor and a specific emotional support one.

As a by-product of the study, we explored whether the HPSS ratings could detect differences between two healthcare providers. In keeping with the literature (Zhou et al., 2011; Whitley, 2012; Ansmann et al., 2014), we expected that patient ratings of perceived social support on the HPSS would differ between two oncology centers reputed for religious and government-operated organizational culture, respectively. Last,

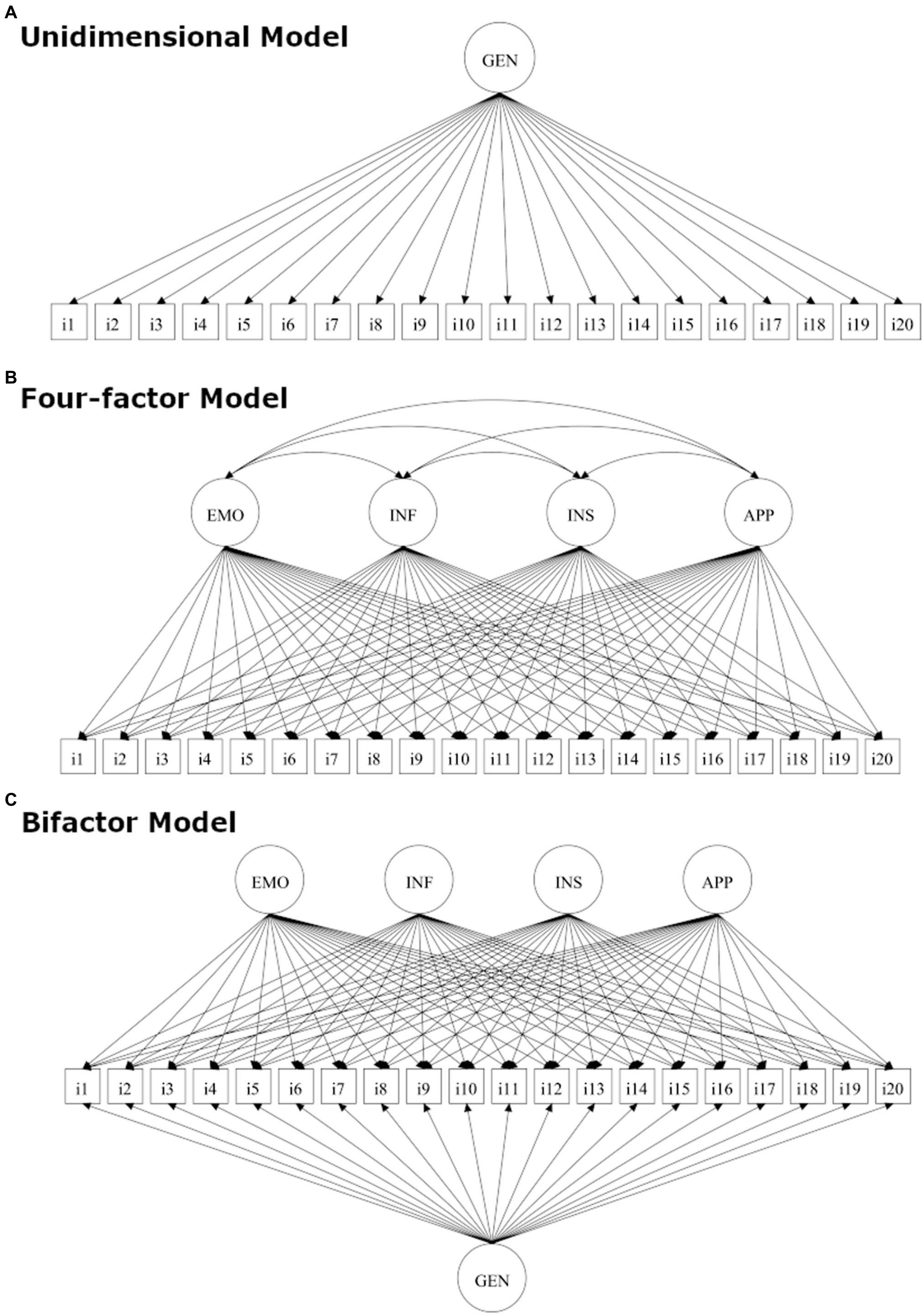

FIGURE 1 | Schematic representations of the models tested in the present study: **(A)** unidimensional model; **(B)** four-factor model; and **(C)** bifactor model. EMO, emotional support; INF, informational support; INS, instrumental support; APP, appraisal support; and GEN, general support.

we explored whether the HPSS was sensitive to how patients at different stages of disease perceived their healthcare provider as providing support for all types or just for specific kinds. This hypothesis was motivated by studies showing that a patient's need for support from a healthcare provider may change during the cancer experience. For instance, late in the course of cancer, especially when patients perceive a poorer prognosis, there is a greater need for emotional and instrumental support from health professionals (e.g., Arora and Gustafson, 2009).

Materials and Methods

Participants and Procedure

One-hundred-sixty-two consecutive patients were recruited from two oncology centers in Rome, Italy, hereafter referred to as hospitals R and G, and surveyed for this study. R was a reputed religious hospital receiving public funding; G was a government-operated hospital. All participants were patients with a confirmed cancer diagnosis receiving chemotherapy in day-treatment units. Inclusion criteria for the study were: a performance status (ECOG) of 0 or 1, age over 18 years old, written comprehension of the Italian language, and ability to fill in a paper and pencil questionnaire. Exclusion criteria were a refusal to cooperate and mental disorders due to medical conditions. The refusal rate was around 5% in both hospitals. No cases were excluded for secondary mental disorders. The ethical review board of Hospital R approved all aspects of this study. Participation was voluntary, and patients had the right to withdraw from the study. Consent was obtained before data collection. Patient characteristics are shown in **Table 1**.

TABLE 1 | Patient characteristics (Study 1 and 2).

Characteristic	Study 1		Study 2	
Age	*M*	*SD*	*M*	*SD*
	58.97	13.28	59.28	12.59
Gender	*N*	(%)	*N*	(%)
Female	89	54.9	40	57.9
Male	73	45.1	29	42.1
Total	162	100	69	100
Tumor site	*N*	(%)	*N*	(%)
Stomac, colon, rectal	60	37.0	15	21.7
Female genitals	22	13.6	–	–
Breast	25	15.4	21	30.4
Skin	–	–	8	11.6
Lung	27	16.7	12	17.4
Kidney, bladder	12	7.4	1	1.4
Male genitals	8	4.9	4	5.8
Other	8	4.9	8	11.6
Total	162	100	69	100
Stage	*N*	(%)	*N*	(%)
I	9	5.6	31	44.9
II	16	9.9	17	24.6
III	36	22.2	6	8.7
IV	101	62.3	15	21.7
Total	162	100	162	100
Hospital	*N*	(%)	*N*	(%)
Government operated	60	37	69	100
Religious	102	63	–	–
Total	162	100	69	100

Measures

Healthcare Provider Social Support Scale

The HPSS administered in the present study consisted of 20 items arranged in four domains selected according to how social support functions have been operationally defined in existing scales used for informal social networks (e.g., Cutrona and Russell, 1987). The item format included the main statement (e.g., "the doctor was listening to you when you talked about your feelings"), followed by one or more illustrative examples (e.g., "pausing to talk to you or beyond the time strictly necessary to carry out his work; or leaving you time to talk to him about your fears"). The authors of this article developed the descriptive statements, which were reviewed with a medical oncologist in one of the two hospitals mentioned above. Thus, domain selection and statement generation followed what Magasi et al. (2012) considered an "etic" approach to content validity, i.e., how clinicians, researchers, or subject matter experts view the concepts to be measured. The illustrative examples were obtained from qualitative interviews with patients conducted by a hospital psychologist. Accordingly, these examples reflected an "emic" approach, i.e., they capitalized on insiders' first-hand experience of the concept being measured (Magasi et al., 2012). The English translation of the HPSS, scoring instructions, and preliminary reference data are reported in Supplementary Materials. Patients were asked to rate their doctors' social support behaviors during the visits or when they stay in the hospital using a five-point frequency scale (1 = Never; 2 = Rarely; 3 = Sometimes; 4 = Often; 5 = Always). Patients were instructed to resort to the examples if the main statement was unclear. Before the study, the HPSS was piloted with a small sample of patients to ensure the understandability of item content, examples, response format, and instructions. No particular problems were encountered in the pilot study.

Data Analysis

Descriptive Analysis

Before ESEM, we assessed item descriptive statistics and item-rest correlations (IRC). In classical test theory, IRC is an index of item discrimination, indicating the extent to which an item separates individuals with high and low scores on the total scale scores. IRCs >0.50, 0.30, and 0.10 are considered strong, moderate, and weak discrimination, respectively (Bechger et al., 2003). Distributional assumptions were checked using the MVN package for R (Korkmaz et al., 2019). Because Shapiro–Wilk's test is sensitive to small departures from univariate normality, it was used in conjunction with an established rule of thumb. Accordingly, we interpreted univariate skewness and kurtosis following Kline (2015), with critical values set at 3 and 10, respectively.

Missing Data

One-hundred-fifty-eight patients (98%) were complete cases. Sporadic missing data were observed. The missing data pattern did not reveal configurations that would indicate missingness not at random (Little's MCAR test = 59.41, df = 61, $p = 0.534$). Therefore, we imputed missing data based on the SPSS Expectation–Maximization procedure. There were no discernible differences between complete-case and imputed data set analyses that would invalidate the study's conclusions. However, in ESEM analysis, we used the imputed dataset to maximize the sample size.

Structural Equation Modeling

ESEM analyses were conducted using *Mplus* (Version 8.4). We compared a bifactor model with the one-factor and four-factor models (**Figure 1**). A bifactor model is a latent structure in which each item loads on a general factor common to all items and a group factor common to some items. The general factor represents the target domain that one is most interested in (e.g., social support by a healthcare provider). The group factors represent narrow factors that explain item responses not accounted for by the general factor (e.g., information provision). According to Marsh et al. (2014), we used a Target Rotation because it provides a robust *a priori* model, allows more control over the model's specification, and makes it easier to interpret the results. Because the HPSS uses Likert-type items, we carried out the analysis using robust least squares estimators (DWLS). This method is recommended to handle ordinal categorical data and has no distributional assumptions (Rhemtulla et al., 2012).

The model's fit was assessed using the DWLSχ2 and other descriptive indices (Kline, 2015). CFI and TLI are incremental indices that compare the fit of the factor model to that of a null model, in which all items are assumed to be uncorrelated. CFI and TLI values >0.90 indicate acceptable fit, while values >0.95 indicate a good fit. The RMSEA measures the difference between the reproduced correlation matrix and the population correlation matrix, controlling sampling variability. An RMSEA of 0.05 or less indicates a close fit, and values up to 0.08 represent a reasonable approximation error. The 90% CI point estimate is also commonly reported to indicate the possibility of a close or exact fit.

The standardized factor loading matrix was analyzed using Dueber (2021) bifactor indices calculator package for R. It provides valuable factor and item statistics for addressing scale dimensionality and evaluating the appropriateness of the bifactor model solution. The ECV assesses the proportion of shared variance explained by each factor in the model, either general or group factors. For the general factor, ECV_G supports unidimensionality when values greater than 0.70 are obtained. ECV_S reflects the proportion of shared variance in subscale items explained by each specific factor. At the item level, IECV assesses how variance in each item can be attributed to the related variation in the general factor alone. Item unidimensionality is supported when values greater than 0.80 are obtained (Rodriguez et al., 2016). The Average Relative Parameter Bias (ARPB) cumulatively assesses the difference between an item's loading on the general factor in the bifactor model and the corresponding loading in the unidimensional model. According to Rodriguez et al. (2016), ARPB values less than 0.15 support unidimensionality. RPB can be used to detect items for which substantial multidimensionality exists at the item level.

Mokken Scale Analysis

Using the Mokken package for R (van der Ark, 2007), we carried out the Mokken Scale Analysis (MSA). It is a nonparametric analog of Rasch analysis but has fewer assumptions about the shape of item characteristic curves (ICC). However, MSA assumes *local independence*, namely that participants' responses to one item are independent of responses to other scale items if the underlying latent construct has been partialled out. This assumption was tested using W_1 and W_3 indices (Straat et al., 2016), which flag positive and negative locally dependent item pairs, respectively. Another assumption is *monotonicity*, namely that the probability of endorsing a specific item response category is a monotonically increasing function of the latent trait. This assumption, and the related *non-intersection* of item response curves, was tested through visual inspection of the ICC. After verifying these assumptions, we examined the *Scalability* of the HPSS total score and subscales scores. Scalability is the extent to which individual items in a scale measure the latent characteristic being measured. In MSA, Loevinger's coefficient H test the level of Scalability for each item (H_i) and the entire set of items that form a scale or a subscale (H_j). A scale based on items with high H_i is highly scalable and likely to be unidimensional. H_j values greater than 0.30 are considered acceptable. According to Sijtsma and Molenaar (2002), Hj values in the range between 1.00 and 0.50 indicate strong Scalability and unidimensionality. Moderate and weak scalability ranges are between 0.49 and 0.40 and between 0.39 and 0.30, respectively.

Results

Descriptive Item Analysis

As one can see from **Table 2**, means and SDs were comparable across items within each social support domain, showing item homogeneity. Even though the Shapiro–Wilks test was significant (all p-$s < 0.01$), the univariate skewness and kurtosis were within the acceptable limits of 3 and 10, respectively. The IRC was above the strong discrimination level for all Emotional, Informational, and Instrumental items, while it was close for Appraisal support items (**Table 2**). Overall, the item set proved adequate to be submitted to factor analysis.

Exploratory Structural Equation Modeling

First, we tested the unifactor model, which assumed that all items would measure only a general perception that healthcare was supportive. This hypothesis was rejected ($\chi^2 = 554.22$; df = 170; $p < 0.001$; TLI = 0.816; CFI = 0.835; RMSEA = 0.118; SRMR = 0.127). Next, we tested the four-factor model with correlated factors, which approached the good fit for most indices ($\chi^2 = 148.51$; df = 116; $p < 0.023$; TLI = 0.977; CFI = 0.986; RMSEA = 0.042; SRMR = 0.043). The target loadings were all statistically significant for this model (all p-$s < 0.005$) and were all greater than 0.40 for Emotional ($\lambda_{range} = 0.62–0.85$), Informational ($\lambda_{range} = 0.43–0.75$) and Instrumental support ($\lambda_{range} = 0.47–0.83$; **Table 3**, Panel a). Except for item #16 ($\lambda = 0.30$), the Appraisal factor also loaded on items greater than 0.40. The factor loadings for appraisal support were also more heterogeneous than for Emotional, Informational, and Instrumental support ($\lambda_{range} = 0.30–0.81$).

TABLE 2 | Descriptive statistics and response frequencies for the Healthcare Perceived Social Support items.

Domain	Item stem	Item descriptive statistics					Item response frequencies (%)				
	The doctor …	***M***	***SD***	***Sk***	***K***	***IRC***	**1**	**2**	**3**	**4**	**5**
Emot.	1. Comforted you by physically expressing his affection	3.90	1.31	−1.04	0.01	0.65	10.5	3.1	18.5	22.2	45.7
Emot.	2. Has been listening to you talk about your feelings	3.41	1.55	−0.43	−1.29	0.74	21.0	6.8	19.8	14.8	37.7
Emot.	3. Has shown interest and concern for your well-being	4.28	1.10	−1.51	1.51	0.49	4.3	3.1	14.8	16.0	61.7
Emot.	4. Let you know that he/she understands your mood and concerns	3.28	1.66	−0.29	−1.55	0.59	27.2	6.2	17.3	10.5	38.9
Emot.	5. Was present and heartened you in a stressful situation for you	3.37	1.59	−0.43	−1.36	0.59	24.1	4.9	17.9	16.0	37.0
Info.	6. Suggested a few actions you should take	3.20	1.67	−0.24	−1.60	0.61	29.0	6.8	14.8	13.6	35.8
Info.	7. Gave you useful information to solve your problem	3.18	1.69	−0.21	−1.63	0.64	30.9	4.9	17.3	9.9	37.0
Info.	8. Explained the pros and cons of each option you had to choose from	3.79	1.59	−0.89	−0.87	0.53	19.8	3.1	11.1	10.5	55.6
Info.	9. Made you aware of what was coming	4.48	1.06	−2.20	3.99	0.31	4.9	3.1	4.9	13.6	73.5
Info.	10. Taught you how to do something	2.91	1.73	0.06	−1.71	0.51	39.5	3.1	17.9	6.8	32.7
Instr.	11. Did some activity with you to help distract you	1.51	1.13	2.25	3.94	0.51	79.0	5.6	8.0	0.6	6.8
Instr.	12. Took you to someone who could act	1.67	1.26	1.73	1.61	0.73	74.1	4.9	9.9	3.1	8.0
Instr.	13. Helped you do something that needed to be done	2.94	1.72	0.01	−1.72	0.41	38.3	4.9	13.0	12.3	31.5
Instr.	14. Lent you or gave you something you needed	1.47	1.10	2.35	4.35	0.57	81.5	3.7	7.4	1.2	6.2
Instr.	15. Performed some tasks for you that you could not do for yourself at that time	1.59	1.20	1.93	2.47	0.59	76.5	4.9	9.3	1.9	7.4

(Continued)

TABLE 2 | Continued

Domain	Item stem	Item descriptive statistics					Item response frequencies (%)				
	The doctor …	***M***	***SD***	***Sk***	***K***	***IRC***	**1**	**2**	**3**	**4**	**5**
Appr.	16. Let you know that he/she approves of the way you deal with situations	2.98	1.70	−0.02	−1.69	0.43	35.8	5.6	15.4	11.7	31.5
Appr.	17. Has expressed appreciation or respect for any of your skills or abilities	1.96	1.47	1.14	−0.32	0.39	66.0	3.1	11.7	6.8	12.3
Appr.	18. Considered you a reliable person, who can be trusted	3.94	1.43	−1.05	−0.33	0.49	12.3	5.6	14.2	12.3	55.6
Appr.	19. Treated you as an equal	4.56	0.98	−2.39	4.88	0.30	3.1	4.3	4.3	9.9	78.4
Appr.	20. Let you know that he/she appreciates you as a “person”	3.70	1.53	−0.80	−0.88	0.51	18.5	3.7	14.2	16.7	46.9

Emot., emotional support; Info., informational support; Instr., instrumental support; Appr., appraisal support; M, item mean; SD, item standard deviation; Sk, univariate skewness index; K, univariate kurtosis index; IRC, item-rest correlation; 1, never; 2, rarely; 3, sometimes; 4, often; and 5, always.

Although there were many significant non-target cross-loadings (i.e., 22 out of 60), these were greater than the target loading only for item #16 (**Table 4**, Panel a). Item simplicity (i.e., IFS > 0.70) was supported for most HPSS items. As one can see from **Table 4** (Panel b), the factor correlations were significant, with a moderate effect size. The four-factor model with correlated factors expands on the one-factor model, showing that the perceived healthcare support could be broken down into separate but still related functional components. Moderate factor correlations can be explained by a general domain factor, suggesting a bifactor model approach to represent the structure of the HPSS.

The bifactor model provided a nearly perfect fit ($\chi^2 = 121.71$; df = 100; $p = 0.069$; TLI = 0.982; CFI = 0.991; RMSEA = 0.037; SRMR = 0.039) and was significantly better than the four-factor model ($\Delta\chi^2 = 28.43$; df = 16; $p < 0.05$). As seen in **Table 5**, all items (except #9 and #19) significantly loaded on the general factor, and the coefficients were generally moderate to large ($\lambda_{range} = 0.37–0.80$). Except for Appraisal Support, for which only three target loadings out of five were statistically significant ($\lambda_{range} = 0.52–0.79$), all target loadings identified Emotional ($\lambda_{range} = 0.43–0.65$), Informational ($\lambda_{range} = 0.30–0.65$) and Instrumental support ($\lambda_{range} = 0.24–0.62$) factors (**Table 5**). As in the previous analysis, the significant cross-loadings did not threaten the simplicity of the factor solution. The index of simplicity calculated for the four group factors showed that the items maintained good purity as indicators of Emotional, Informational, Instrumental, and Appraisal support (**Table 4**).

Given that HPSS items entangled general and specific sources of variance, the following analyses were performed to determine whether total or subscale scores are supported. The general factor explained about half of the common variance in perceived social support ($ECV_G = 51\%$) the remaining common variance was explained almost equally by Affective ($ECV_S = 14\%$), Informational ($ECV_S = 12\%$), Instrumental ($ECV_S = 10\%$), and Appraisal ($ECV_S = 13\%$) factors. The model ARPB was equal to 37%. Given $ECV_G < 70\%$ and ARPB >15%, we can conclude that multidimensionality cannot be neglected in modeling the health provider social support construct using the HPSS scale. At the item level, only four items (#10, #11, #16, and #17) with IECV >0.80 varied between participants because of variation in the general factor only. The remaining items entangled to varying degrees a portion of variance related to the general factor and a portion related to the group factor ($IECV_{range} = 0.02–0.67$).

Following Rodriguez et al. (2016), we compared the reliability ω coefficients separately assessed for each subscale to the corresponding hierarchical ones (ωsh) in which the general factor variance was partialled out. This comparison is to assess the unique information provided by each of the subscale scores relative to the total reliable variance in each subscale. The resulting ω coefficients were high: 0.92, 0.90, 0.89, and 0.92 for Emotional, Informational, Appraisal, and Instrumental subscales, respectively. The ωsh dropped to 0.41, 0.39, 0.59, and 0.29, respectively. For the total score, instead, the proportion of reliable variance was $\omega = 0.95$, with 81% of the variance depending on the general domain factor (i.e., $\omega h = 0.81$). So, using the total score would dilute too much the variance of the specific factors present in the subscales.

Mokken Scale Analysis

First, we evaluated the local independence assumption. For the total score (20 items), we found only one flagged item pair (i.e., #2 with #3) to be positively locally dependent; no

TABLE 3 | Standardized factor loadings and factor intercorrelations from the exploratory structural equation modeling four-factor solution of the HPSS.

Item	**Panel a: factor loadings**								
	Emotional		**Informational**		**Instrumental**		**Appraisal**		**IFS**
	λ	*p*	*λ*	*p*	*λ*	*p*	*λ*	*p*	
1	**0.80**	**(0.000)**	0.02	(0.829)	0.00	(0.975)	−0.04	(0.590)	1.00
2	**0.85**	**(0.000)**	0.05	(0.445)	0.08	(0.092)	0.03	(0.543)	0.99
3	**0.62**	**(0.000)**	0.16	(0.037)	−0.33	(0.000)	0.26	(0.000)	0.66
4	**0.66**	**(0.000)**	−0.07	(0.355)	0.10	(0.064)	0.22	(0.003)	0.87
5	**0.65**	**(0.000)**	0.23	(0.000)	0.16	(0.007)	−0.11	(0.097)	0.82
6	0.18	(0.015)	**0.58**	**(0.000)**	0.33	(0.000)	−0.09	(0.236)	0.69
7	−0.05	(0.479)	**0.71**	**(0.000)**	0.37	(0.000)	0.02	(0.840)	0.78
8	−0.06	(0.421)	**0.75**	**(0.000)**	0.11	(0.142)	−0.01	(0.843)	0.97
9	0.13	(0.116)	**0.59**	**(0.000)**	−0.31	(0.001)	0.25	(0.005)	0.66
10	0.31	(0.001)	**0.43**	**(0.000)**	0.28	(0.001)	−0.09	(0.393)	0.50
11	0.34	(0.000)	0.03	(0.738)	**0.64**	**(0.000)**	0.05	(0.630)	0.77
12	−0.13	(0.167)	0.26	(0.001)	**0.83**	**(0.000)**	0.08	(0.307)	0.88
13	−0.03	(0.728)	0.30	(0.000)	**0.47**	**(0.000)**	0.15	(0.119)	0.66
14	0.07	(0.475)	0.04	(0.748)	**0.80**	**(0.000)**	0.09	(0.439)	0.98
15	0.00	(0.983)	0.15	(0.112)	**0.74**	**(0.000)**	0.17	(0.151)	0.91
16	0.44	(0.000)	−0.04	(0.601)	0.31	(0.000)	**0.30**	**(0.004)**	0.24
17	−0.04	(0.667)	−0.12	(0.226)	0.64	(0.000)	**0.42**	**(0.001)**	0.29
18	−0.05	(0.568)	−0.05	(0.540)	−0.04	(0.615)	**0.81**	**(0.000)**	0.99
19	−0.01	(0.959)	0.35	(0.000)	−0.33	(0.000)	**0.68**	**(0.000)**	0.67
20	0.08	(0.385)	−0.19	(0.034)	0.17	(0.051)	**0.72**	**(0.000)**	0.88
	Panel b: factor correlations								
Factor	*ϕ*	*p*	*ϕ*	*p*	*ϕ*	*p*	*ϕ*	*p*	
Emotional									
Informational	0.40	(0.000)							
Instrumental	0.31	(0.000)	0.23	(0.001)					
Appraisal	0.47	(0.000)	0.36	(0.000)	0.14	(0.031)			
	Emotional		Informational		Instrumental		Appraisal		

λ, factor loading; ϕ, factor correlation; p, p-level; IFS, index of factorial simplicity, values in the 0.90s, the 80s, and the 70s are interpreted as marvelous, meritorious, and middling, respectively. Target factor loadings are shown in bold; significant non-target loadings are underlined.

negative local dependence was found. For the Affective and Instrumental subscales (with five items each), we found one flagged item pair to be positively locally dependent (i.e., #2 with #3 and #12 with #13, respectively). No violations were detected for the Informational and Appraisal subscales. Therefore, the local independence assumption was only sporadically violated. Second, we inspected the item characteristic curves reported in Supplementary Materials. All items comprising the total score had monotonically non-decreasing patterns, and no significant violations were detected (**Supplementary Figure S1**; **Supplementary Table S1**). The analysis of subscales revealed some violations of the monotonicity assumptions (**Supplementary Figure S2**); however, only three items (#2, #8, and #18 belonging to Emotional, Information, and Appraisal scores, respectively) yielded statistically significant violations (**Supplementary Table S2**). Taken together, the assumptions for testing item scalability are overall tenable, with a few exceptions.

Table 5 shows the scalability coefficients H_i for specific items under total and subscale scoring approaches. The global H_j for the total scale and separate subscales were also reported. Using the total score approach, five items out of 20 were poorly scalable, and the resulting scale was weakly scalable. Conversely, scoring the HPSS according to a subscale approach made most items moderately to highly scalable. The Emotional and Instrumental HPSS subscales were strong, while the Informational and Appraisal ones were moderate. The Mokken reliability coefficients were 0.88 for the total score and 0.83, 0.77, 0.83, and 0.69 for the Emotional, Informational, Instrumental, and Appraisal subscales.

Preliminary Criterion Validity Analyses

One of the intended applications of the HPSS could be assessing the perceived level of social support provided by healthcare providers in different hospitals, outpatient clinics, services, or departments in studies of healthcare quality. We piloted this approach, involving hospitals R and G in the HPSS preliminary validation study. Patients were also stratified by disease stage (i.e., stages I–III vs. stage IV) in data analysis because the patient's need for support from a healthcare provider may change during the cancer experience. The analysis of Emotional, Informational, Instrumental, Appraisal support by hospital and disease stage revealed statistically significant multivariate main effects for the hospital ($F = 20.83$; df = 4,157; $p < 0.001$) and hospital × support function interaction ($F = 6.39$; df = 3,155; $p < 0.001$) accounting for 12 and 11% of the variance in the combined dependent variables, respectively. Follow-up univariate analyses were examined to determine which functional components of social support accounted for the multivariate effect. Significance was determined at $p < 0.0125$ (i.e., $\alpha = 0.05/4$) to control for familywise type 1 error. Group means are shown

TABLE 4 | Standardized factor loadings and factor intercorrelations from the exploratory structural equation modeling bifactor solution of the HPSS.

	Factor loadings										Item indexes		
	General		Emotional		Informational		Instrumental		Appraisal				
	λ	*p*	λ	*p*	λ	*p*	λ	*p*	λ	*p*	IECV	RPB	IFSs
1	**0.56**	**(0.000)**	**0.55**	**(0.000)**	0.01	(0.872)	−0.12	(0.112)	−0.01	(0.840)	0.50	0.22	0.96
2	**0.68**	**(0.000)**	**0.62**	**(0.000)**	0.02	(0.700)	0.00	(0.959)	0.04	(0.418)	0.54	0.23	0.99
3	**0.33**	**(0.002)**	**0.65**	**(0.000)**	0.16	(0.038)	−0.08	(0.386)	0.34	(0.000)	0.16	0.87	0.74
4	**0.62**	**(0.000)**	**0.45**	**(0.000)**	−0.07	(0.299)	−0.05	(0.538)	0.16	(0.015)	0.61	0.15	0.86
5	**0.64**	**(0.000)**	**0.43**	**(0.000)**	0.17	(0.004)	0.02	(0.845)	−0.08	(0.129)	0.66	0.14	0.84
6	**0.64**	**(0.000)**	0.06	(0.338)	**0.43**	**(0.000)**	0.10	(0.288)	−0.10	(0.145)	0.67	0.09	0.89
7	**0.68**	**(0.000)**	−0.19	(0.004)	**0.58**	**(0.000)**	0.01	(0.906)	−0.03	(0.725)	0.56	0.02	0.90
8	**0.37**	**(0.000)**	0.06	(0.393)	**0.65**	**(0.000)**	0.19	(0.025)	0.05	(0.569)	0.23	0.44	0.91
9	**0.22**	**(0.092)**	0.33	(0.000)	**0.55**	**(0.000)**	−0.08	(0.521)	0.35	(0.000)	0.08	1.25	0.56
10	**0.67**	**(0.000)**	0.09	(0.222)	**0.30**	**(0.000)**	−0.08	(0.453)	−0.13	(0.121)	0.78	0.01	0.75
11	**0.80**	**(0.000)**	0.06	(0.451)	−0.07	(0.471)	**0.24**	**(0.032)**	−0.10	(0.408)	0.89	0.03	0.77
12	**0.63**	**(0.000)**	−0.13	(0.062)	0.14	(0.042)	**0.68**	**(0.000)**	−0.05	(0.480)	0.45	0.16	0.92
13	**0.54**	**(0.000)**	−0.04	(0.603)	0.21	(0.002)	**0.34**	**(0.001)**	0.08	(0.353)	0.63	0.07	0.68
14	**0.68**	**(0.000)**	−0.04	(0.541)	−0.08	(0.444)	**0.52**	**(0.000)**	−0.08	(0.472)	0.62	0.11	0.95
15	**0.67**	**(0.000)**	−0.03	(0.765)	0.03	(0.720)	**0.51**	**(0.000)**	0.01	(0.906)	0.63	0.12	0.99
16	**0.74**	**(0.000)**	0.18	(0.026)	−0.09	(0.199)	−0.07	(0.324)	**0.15**	**(0.183)**	0.89	0.06	0.33
17	**0.70**	**(0.000)**	−0.25	(0.002)	−0.21	(0.026)	0.17	(0.068)	**0.18**	**(0.221)**	0.74	0.20	0.19
18	**0.37**	**(0.002)**	−0.01	(0.871)	−0.02	(0.843)	−0.24	(0.011)	**0.71**	**(0.000)**	0.20	0.19	0.89
19	**0.13**	**(0.315)**	0.34	(0.000)	0.35	(0.000)	0.13	(0.153)	**0.79**	**(0.000)**	0.02	3.02	0.71
20	**0.49**	**(0.000)**	0.07	(0.446)	−0.16	(0.054)	−0.02	(0.858)	**0.52**	**(0.000)**	0.43	0.04	0.90

λ, factor loading; p, p-level. IECV, individual item explained common variance, items with large loadings on the general factor and IECV greater than 0.80 or 0.85 will typically yield a unidimensional item set that reflects the content of the general domain; RPB, individual item relative parameter bias, parameter bias less than 0.10–0.15 is acceptable and poses no serious concern regarding multidimensionality. IFSs, index of factorial simplicity for group-specific factors, values in the 0.90s, the 80s, and the 70s are interpreted as marvelous, meritorious, and middling, respectively. Target factor loadings are shown in bold; Significant non-target loadings are underlined.

TABLE 5 | Descriptive statistics and intercorrelations among HPSS scores.

(a) Analysis of total score					(b) Analysis of separate subscale scores				
Item	***Hi***	***SE***	***Z*-score**		**Item**	***Hi***	***SE***	***Z*-score**	
1	0.36	(0.04)	8.26		1	0.55	(0.05)	10.86	
2	0.42	(0.04)	11.69		2	0.59	(0.04)	14.49	
3	0.30	(0.06)	5.43		3	0.46	(0.07)	6.71	
4	0.36	(0.04)	8.62		4	0.50	(0.05)	9.33	
5	0.39	(0.04)	9.63		5	0.50	(0.05)	9.28	Hj = 0.53
6	0.39	(0.04)	10.18		6	0.49	(0.05)	9.35	
7	0.37	(0.04)	9.35		7	0.51	(0.05)	10.63	
8	0.29	(0.05)	6.06		8	0.45	(0.05)	8.39	
9	0.28	(0.06)	4.44		9	0.36	(0.08)	4.31	
10	0.37	(0.04)	8.74		10	0.45	(0.06)	7.50	Hj = 0.46
11	0.55	(0.07)	8.07		11	0.46	(0.09)	5.24	
12	0.47	(0.07)	7.17		12	0.61	(0.06)	10.72	
13	0.32	(0.05)	6.49		13	0.56	(0.07)	7.53	
14	0.50	(0.09)	5.91		14	0.52	(0.10)	5.52	
15	0.48	(0.07)	6.59		15	0.52	(0.08)	6.54	Hj = 0.54
16	0.40	(0.04)	9.88		16	0.38	(0.06)	6.62	
17	0.37	(0.07)	5.44		17	0.49	(0.08)	6.16	
18	0.23	(0.05)	5.02		18	0.42	(0.05)	8.63	
19	0.29	(0.08)	3.91		19	0.35	(0.07)	4.69	
20	0.27	(0.05)	6.00	Hj = 0.36	20	0.43	(0.05)	7.89	Hj = 0.42

Hi, scalability coefficient for individual items; SE, standard error of Hi; and Hj, scalability coefficient for the total score or subscale scores.

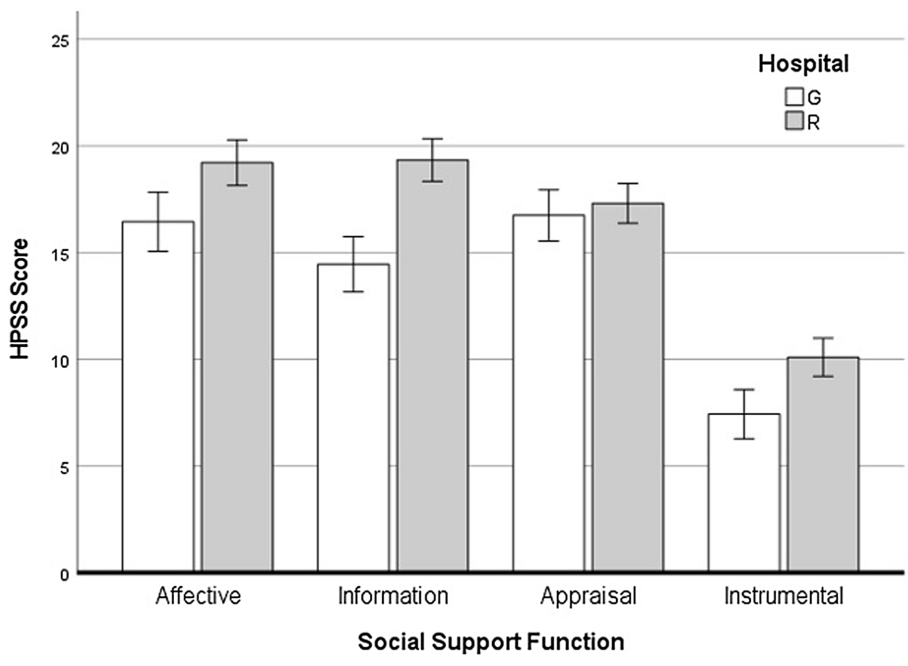

FIGURE 2 | Differences between government-operated (G) and religious (R) hospitals across different social support functions.

in **Figure 2**. Patients receiving treatment in the religious hospital felt more supported than those treated in the public hospital from an emotional, informative, and practical point of view. There was no difference in esteem support. Using the HPSS total score as the dependent variable, we found that patients at hospital R perceived their healthcare provider to be overall more supportive.

STUDY 2: HPSS AND RELATED CONSTRUCTS

The previous study showed that the HPSS scale has a solid factorial structure and that the four social support functions can be reliably measured. However, examining only the factorial structure of the scale is not sufficient to demonstrate the validity of HPSS scores. Given that validation of a new measure is a laborious process that requires a range of empirical evidence, in this study, we surveyed an independent sample of oncology patients to take the first step in this direction. As part of criterion-related validity assessment, we aimed to establish the correlations between HPSS scores and two critical variables in the physician-patient relationship. These were the physician's communication skills and the patient's trust toward the healthcare provider. Recent studies have reinforced the view that patients of health professionals with better communication skills develop trusting relationships with their doctors (Chandra et al., 2018) and have better health outcomes (Howick et al., 2018; Noble, 2020). Accordingly, we expected the HPSS scores to positively correlate with the doctor's communication skills and the patient's trust toward the health care provider.

Materials and Methods

Participants and Procedure

Sixty-nine non-consecutive patients were recruited from one oncology center in Naples, Italy. As in Study 1, all participants had a confirmed cancer diagnosis and received chemotherapy in a day-treatment unit. Inclusion and exclusion criteria were the same as in Study 1. The refusal rate was around 10%, and no cases were excluded because of secondary mental disorders. The ethical review board at Sapienza University in Rome approved the study. As a condition of participation, all patients provided their informed consent. Characteristics of patients can be found in **Table 1**.

Measures

Healthcare Provider Social Support Scale

Same as in Study 1.

Trust in the Physician Scale

This 11-item scale (TPS; Anderson and Dedrick, 1990) is one of the most widely used tools to assess patients' trust in their physician (Müller et al., 2014). In the absence of a formal validation study for this scale in Italian, the second author of this paper translated items and instructions for use in the present study, receiving back-translation feedback from an expert bilingual professional translator. The TPS uses a five-point response scale (1 = Strongly Disagree to 5 = Strongly Agree) and yields a total score reflecting greater trust.

Communication Assessment Tool

This scale consists of 15 items designed to assess patient perceptions of physician interpersonal and communication skills (CAT; Makoul et al., 2007). The Italian translation of the CAT has been recently tested in an outpatient surgical clinic (Scala et al., 2016). Patients are asked to answer each item based on a single, recent physician interaction. The CAT yields a summary score reflecting the patient's overall satisfaction.

Data Analysis

Pearson's r was used to assess the correlations among variables. Nonlinear correlations were explored using the nlcor R package (Ranjan and Najari, 2022). The unique contribution of social support and communication scores in predicting trust was explored using linear regression analyses. The sequence of analyses and the choice of independent and dependent variables was guided by emerging findings in the correlation analyses and the reviewed literature. As in Study 1, sporadic missing data were observed (59 patients, 86%, were complete cases), and the missing data pattern was completely random (Little's MCAR test = 472.14, df = 471, $p = 0.477$). Due to the relatively small sample size, missing data were imputed, as in Study 1.

Results

Table 6 reports descriptive statistics, reliability coefficients, and correlations among HPSS scores, Trust in Physician, and Physician's Communication Skills. Nonlinear correlations were virtually identical to Pearson correlations, supporting the linearity and monotonicity of all relationships. All the coefficients were positive and statistically significant. The correlations among HPSS scores were very high, reflecting a substantial proportion of shared common variance. Emotional and Informational support functions were more strongly associated with Physician's Communication Skills than Instrumental and Appraisal ones. The total HPSS score was also strongly associated with the communication score. This result showed that physicians with better communication skills were more effective in providing social support to their patients, making them feel secure (i.e., Emotional support) and informed about their treatment and medical status (i.e., Informational support).

Similarly, Emotional, Informational, and Appraisal support functions and the HPSS total score were more strongly associated with Trust in Physician than Instrumental support. Physician's Communication Skills were also associated with the perceived trust, but the correlation was slightly lower than those assessed between trust and social support functions. These findings suggested that healthcare provider social support and doctor-patient communication are fundamental to developing a trusting relationship and therapeutic alliance with their doctors.

Although correlations in a cross-sectional study cannot prove causal relationships, the data collected are consistent with the view that Affective, Informational, Instrumental social support, and Physician's Communication Skills could predict Trust to a different extent and better than instrumental support. We performed linear regression analyses to explore this possibility and disentangle unique social support and communication contributions. We used the HPSS scores as predictors of trust, controlling for Physician's Communication Skills. These analyses showed that Emotional (Beta = 0.46; t-value = 3.02; $p = 0.004$), Informational (Beta = 0.50; t-value = 3.33; $p = 0.001$), and the HPSS total score (Beta = 0.59; t-value = 3.79; $p < 0.001$) remained statistically associated with Trust, making Physician's Communication skills no longer significant. Both Appraisal support (Beta = 0.39; t-value = 2.89; $p = 0.005$) and Physician's Communication skills (Beta = 0.32; t-value = 2.38; $p = 0.021$) uniquely predicted Trust scores. Confirming the correlations reported In **Table 6**, Instrumental support did not predict Trust score, and Physician's Communication was the only significant predictor (Beta = 0.46; *t*-value = 3.80; $p < 0.001$). Overall, these analyses showed that Emotional and Informational healthcare supports were essential to building trusting relationships above and beyond Physician's Communication skills.

DISCUSSION

In the present study, we proposed a new scale, the HPSS, to assess healthcare social support in oncology settings. The scale was designed according to a multidimensional approach, as established in psychosocial research where social networks perform Emotional, Informational, Appraisal, and Instrumental functions (Cutrona and Russell, 1987; Langford et al., 1997; Thoits, 2011; Wang et al., 2014).

Although still preliminary, the results obtained from two studies of patients recruited from different oncology day-treatment units are promising and can be summarized as follows. First, the HPSS had a multidimensional structure. Second, the HPSS subscales proved reliable and preserved a good degree of specific information. Third, we detected differences in perceived social support between patients admitted to religious and government-operated hospitals. Fourth, the Affective, Informational, and Appraisal functions were positively correlated with doctor communication skills and patient's trust in the physician.

Regarding the factor structure, our study showed that the four-factor model had a good fit to the data, outperforming the unifactorial model. Other studies measured healthcare provider social support either as Emotional or as Informational, implicitly assuming that these functions were enough to cover the content domain and precluding comprehensive tests of construct dimensionality (Kuuppelomäki, 2003; Wenrich et al., 2003; Rutten et al., 2005; Ansmann et al., 2012). Our findings indicated that oncology patients could discriminate how a healthcare provider could support them beyond merely demonstrating love and caring, encouragement, and empathy (Thoits, 2011) or providing facts or advice (Wang et al., 2014). Instrumental and Appraisal emerged as distinct social support factors. In a doctor-patient relationship, it could be argued that providing practical, tangible help to patients (Langford et al., 1997; Wang et al., 2014) or endorsing the appropriateness of acts or statements made by patients (Langford et al., 1997) is not required. Nevertheless, even assuming this is true for all medical specialties, it is still possible that other health professionals (e.g., nurses or physical therapists) might contribute to enhancing the perceived quality of care by providing complementary functions of social support. Foreshadowing future research, one could investigate whether Instrumental and Appraisal items apply equally well to physicians and other members of a multidisciplinary oncology care team.

Our study also revealed that a general factor coexisted with the abovementioned factors. CFA has recently complained about necessitating insubstantial or inflated general factors to achieve an acceptable fit (Morin et al., 2016; Joshanloo et al., 2017). Because we used an ESEM approach, which does not produce such statistical artifacts, we are reasonably confident that the general factor represented an additional source of reliable variance. However, the general factor could be challenging to explain. Alternative interpretations are possible. On the one hand, the general factor might reflect a "true" perception that accounts for patients' overall feelings of being supported by healthcare professionals. This interpretation is consistent with previous research considering healthcare social support as a single evaluative dimension, disregarding fine-grained distinctions between specific functions (Katz et al., 2003; Reynolds and Perrin, 2004). On the other hand, the general factor could reflect acquiescence, social desirability, or other response sets. We believe that the first interpretation is the most likely. However, we cannot rule out the second interpretation based on the present study. Thus, future research should validate the general factor against independent response-bias measures.

Although our study supported the construct validity of the HPSS, a few items showed low loadings on the target

TABLE 6 | Descriptive statistics and intercorrelations among HPSS scores, trust in physician, and physician's communication skills.

Score (range)	*α*	*M*	*SD*	Correlations						
Emotional (1–25)	0.92	13.17	5.92	--	*0.75***	*0.75***	*0.74***	*0.91***	*0.64***	*0.76***
Informational (1–25)	0.88	16.06	5.47	0.75**	--	*0.68***	*0.74***	*0.89***	*0.66***	*0.74***
Instrumental (1–25)	0.78	10.12	4.70	0.76**	0.68**	--	*0.72***	*0.87***	*0.50***	*0.57***
Appraisal (1–25)	0.93	14.99	6.21	0.74**	0.74**	0.72**	--	*0.90***	*0.61***	*0.69***
HPSS Total (1–100)	0.96	54.33	19.96	0.91**	0.89**	0.87**	0.90**	--	*0.68***	*0.78***
TPS (15–55)	0.93	39.88	11.33	0.65**	0.66**	0.51**	0.62**	0.69**	--	*0.60***
CAT (20–75)	0.96	51.10	15.05	0.76**	0.74**	0.57**	0.69**	0.78**	0.60**	--
				Emotional	Informational	Instrumental	Appraisal	Total	TPS	CAT

*Pearson's correlations reported. HPSS, health provider social support; TPS, trust in physician scale; CAT, communication assessment tool. Pearson correlations are reported below the diagonal; nonlinear correlations (italicized) are above the diagonal. N = 69. **p < 0.01 (two-tailed).*

factors or lacked factorial simplicity in ESEM analysis; a few others were shown to violate Mokken scaling assumptions. As noted elsewhere (e.g., Morin et al., 2015), the inherent difficulty in producing items that perfectly reflect the constructs intended to measure could explain these defects. However, we cannot exclude other possibilities. For example, in developing the HPSS, we adopted a mostly etic approach to content validity and item generation. However, the items also reflected an emic view of patients in the illustrative examples provided below each item (see **Supplementary Materials**). While combining etic/emic approaches is desirable to widen the content coverage (Magasi et al., 2012), the specific item format (i.e., general statement + illustrative examples) might have increased the item complexity. Of course, defective items might be dropped out from a revision of the HPSS; however, we did not find severely biased or insubstantial items in the present study. So, we think modifying items could be more appropriate than a deletion. Notably, notwithstanding imperfections, the subscale scores preserved a non-negligible amount of reliable information and were reasonably scalable. Therefore, we recommend using subscales in clinical assessment and research applications.

The present study also made some steps toward a more robust assessment of criterion-related and concurrent validity of the HPSS scores. In keeping with previous research (White and Begun, 1998; Reinikka and Svensson, 2010; White et al., 2010; Zhou et al., 2011; Whitley, 2012; Calegari et al., 2015; Seemann et al., 2015), we expected the HPSS to be sensitive to differences in perceived support provided to patients admitted to religious or government-operated hospitals. This hypothesis was supported for Emotional, Informational, and Instrumental support, perceived higher by patients in the religious hospital. This finding implies that the HPSS could be used in quality of care research to assess and compare hospitals, outpatient clinics, particular services, or departments. As part of criterion-related validity, we also expected stage-related differences in perceived social support as suggested by previous research (Arora and Gustafson, 2009). This hypothesis was not confirmed, however. Either the patients have not changed their demands on the doctors, or the doctors may not have adapted to the patient's changing needs. A longitudinal study would be needed to address this issue.

Construct validation is a long-term endeavor, requiring multiple studies and accumulating evidence for the instrument's validity. Our study aimed to establish the correlations between HPSS scores and two critical variables in the physician-patient relationship as part of this process. Constructs similar to social support functions have been used in medical research (Donabedian, 2005; Brédart et al., 2005a,b; Malley and Fernández, 2010; Müller et al., 2014). From our perspective, a physician's communication skills and trustworthiness are not social support variables; nevertheless, we expected to see positive correlations between HPSS scores and these variables. Overall, our hypotheses about the relationships of HPSS ratings with similar constructs in the nomological network were confirmed.

Beyond merely providing initial evidence of HPSS concurrent validity, the correlation pattern reported in the present study was compatible with the view that healthcare provider social support was needed to develop a trusting doctor-patient relationship, above and beyond the physician's excellent communication abilities. Indeed, Affective and Informational scores remained statistically associated with patients' trust in their physician, controlling for the physician's communication skills. Not only were the patients of doctors with excellent communication skills more likely to develop trustworthy relationships (Chandra et al., 2018), but specific healthcare support functions can also have a unique role in increasing patients' trust.

Limitations and Future Directions

The present study is not exempt from limitations. First, the sample used was non-probabilistic and relatively small for structural equation modeling. Nevertheless, we surveyed consecutive clinical patients from different oncology units with a high response rate. So, our sample reflects the typical user of these health services. At the same time, the result of our study might not apply to other clinical populations suffering from different diseases or psychological disorders. Although no golden rule exists for establishing the minimum sample size for structural equation modeling (Kline, 2015), analyses carried out on small samples may fail to converge or provide an improper solution. In our study, we did not encounter any of these problems; therefore, we believe the sample size was sufficient, at least from a computational point of view. Of course, future studies should consider cross-validation of the factor structure on a larger sample. Second, the factors emerging from structural equation modeling need further validation. In particular, the "true" nature of the general factor needs to be clarified. To this purpose, a new data collection must include external measures of social desirability and acquiescence to rule out the possibility that the general factor captured primarily response set variance. Third, the sample size of Study 2 was somewhat limited, and the research design was cross-sectional. Therefore, caution must be exercised to avoid overgeneralizing the results and speculating on possible causal relationships between the variables involved. More research is needed with larger samples and longitudinal designs to rule out alternative interpretations of correlational evidence. Last, all variables in the study are self-reported. This characteristic of our research might have inflated the observed correlations between HPSS scores and the criteria used in the second study. Future validation studies of healthcare social support functions are needed and should compare self-reported data to data extracted from medical records or clinical test results.

Conclusion

Our study showed that patients can discriminate well between different ways healthcare providers can support them, and the scale proposed here can measure healthcare support as a multidimensional construct. Because healthcare provider social support can soothe psychological distress resulting from a poor adjustment to chronic conditions, a multidimensional scale might help to profile which type of social support is more salient in particular healthcare services, and which is lacking. Further longitudinal studies are needed to clarify the reciprocal relationships between social support, physician–patient communication, trust, quality of care, and health outcomes.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Studies involving human participants were reviewed and approved by the Hospital R Ethics Committee (Study 1) and the Ethics Review Committee for Psychological Research, Department of Social and Developmental Psychology, Sapienza University of Rome (Study 2). The Hospital R ethics committee waived the requirement for written informed consent for participation in Study 1. We obtained written informed consent from all patients participating in Study 2 as per the Ethics Review Committee for Psychological Research recommendations.

AUTHOR CONTRIBUTIONS

MT and ML contributed equally to the theoretical and empirical aspects of the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Sapienza Institutional Research grants (RP1181642870C00B, RM11916B7E07E816).

ACKNOWLEDGMENTS

The authors wish to thank Valentina Romeo and Cristina Soldati (Study 1), Ludovica Maria Ercolessi and Romeo Miraglia (Study 2) for help in data collection, and all medical professionals, nurses, and healthcare teams involved in the study. The authors also express sincere gratitude to all patients and their caregivers for participating in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.773447/full#supplementary-material

REFERENCES

Anderson, L. A., and Dedrick, R. F. (1990). Development of the trust in physician scale: a measure to assess interpersonal trust in patient-physician relationships. *Psychol. Rep.* 67, 1091–1100. doi: 10.2466/pr0.1990.67.3f.1091

Ansmann, L., Kowalski, C., Ernstmann, N., Ommen, O., and Pfaff, H. (2012). Patients' perceived support from physicians and the role of hospital characteristics. *Int. J. Qual. Health Care* 24, 501–508. doi: 10.1093/intqhc/mzs048

Ansmann, L., Wirtz, M., Kowalski, C., Pfaff, H., Visser, A., and Ernstmann, N. (2014). The impact of the hospital work environment on social support from physicians in breast cancer care. *Patient Educ. Couns.* 96, 352–360. doi: 10.1016/j.pec.2014.07.016

Arora, N. K., and Gustafson, D. H. (2009). Perceived helpfulness of physicians' communication behavior and breast cancer patients' level of trust over time. *J. Gen. Intern. Med.* 24, 252–255. doi: 10.1007/s11606-008-0880-x

Asparouhov, T., and Muthén, B. (2009). Exploratory structural equation modeling. *Struct. Equ. Model.* 16, 397–438. doi: 10.1080/10705510903008204

Ban, Y., Li, M., Yu, M., and Wu, H. (2021). The effect of fear of progression on quality of life among breast cancer patients: the mediating role of social support. *Health Qual. Life Outcomes* 19:178. doi: 10.1186/s12955-021-01816-7

Barrera, M., and Ainlay, S. L. (1983). The structure of social support: a conceptual and empirical analysis. *J. Community Psychol.* 11, 133–143. doi: 10.1002/1520-6629(198304)11:2<133::AID-JCOP2290110207>3.0.CO;2-L

Beattie, M., Murphy, D. J., Atherton, I., and Lauder, W. (2015). Instruments to measure patient experience of healthcare quality in hospitals: a systematic review. *Syst. Rev.* 4:97. doi: 10.1186/s13643-015-0089-0

Bechger, T. M., Maris, G., Verstralen, H. H. F. M., and Béguin, A. A. (2003). Using classical test theory in combination with item response theory. *Appl. Psychol. Meas.* 27, 319–334. doi: 10.1177/0146621603257518

Brédart, A., Bottomley, A., Blazeby, J. M., Conroy, T., Coens, C., D'Haese, S., et al. (2005a). An international prospective study of the EORTC cancer in-patient satisfaction with care measure (EORTC IN-PATSAT32). *Eur. J. Cancer* 41, 2120–2131. doi: 10.1016/j.ejca.2005.04.041

Brédart, A., Bouleuc, C., and Dolbeault, S. (2005b). Doctor-patient communication and satisfaction with care in oncology. In *Current Opinion in Oncology* 17, 351–354).

Breuer, N., Sender, A., Daneck, L., Mentschke, L., Leuteritz, K., Friedrich, M., et al. (2017). How do young adults with cancer perceive social support? A qualitative study. *J. Psychosoc. Oncol.* 35, 292–308. doi: 10.1080/07347332.2017.1289290

Bucholz, E. M., Strait, K. M., Dreyer, R. P., Geda, M., Spatz, E. S., Bueno, H., et al. (2014). Effect of low perceived social support on health outcomes in young patients with acute myocardial infarction: results from the VIRGO (variation in recovery: role of gender on outcomes of young AMI patients) study. *J. Am. Heart Assoc.* 3:e001252. doi: 10.1161/JAHA.114.001252

Calegari, R. C., Massarollo, M. C., and Santos, M. J. (2015). Humanização da assistência à saúde na percepção de enfermeiros e médicos de um hospital privado. *Rev. Esc. Enferm. USP* 49, 42–47. doi: 10.1590/S0080-623420150000800006

Chandra, S., Mohammadnezhad, M., and Ward, P. (2018). Trust and communication in a doctor-patient relationship: a literature review. *J. Health Commun.* 3:36. doi: 10.4172/2472-1654.100146

Commission on Social Determinants of Health (2008). Closing the gap in a generation: health equity through action on the socialdeterminants of health. *World Health Organ.* 6, 102–105. doi: 10.1080/17441692.2010.514617

Cutrona, C. E., and Russell, D. W. (1987). "The provisions of social relationships and adaptation to stress" in *Advanced Personal Relationships (Issue October)*. eds. D. Perlman and W. Jones (Greenwich, CT: JAI Press)

Donabedian, A. (2005). Evaluating the quality of medical care. *Milbank Q.* 83, 691–729. doi: 10.1111/j.1468-0009.2005.00397.x

Dueber, D. (2021). BifactorIndicesCalculator: Bifactor Indices Calculator. R package version 0.2.2. Retrieved from https://CRAN.R-project.org/package=BifactorIndicesCalculator

Fleming, G. V. (1981). Hospital structure and consumer satisfaction. *Health Serv. Res.* 16, 43–63. PMID: 7228714

Graven, L. J., and Grant, J. (2013). The impact of social support on depressive symptoms in individuals with heart failure: update and review. *J. Cardiovasc. Nurs.* 28, 429–443. doi: 10.1097/JCN.0b013e3182578b9d

Graven, L. J., and Grant, J. S. (2014). Social support and self-care behaviors in individuals with heart failure: an integrative review. *Int. J. Nurs. Stud.* 51, 320–333. doi: 10.1016/j.ijnurstu.2013.06.013

Hancock, S. L., Ryan, O. F., Marion, V., Kramer, S., Kelly, P., Breen, S., et al. (2020). Feedback of patient-reported outcomes to healthcare professionals for comparing health service performance: a scoping review. *BMJ Open* 10:e038190. doi: 10.1136/bmjopen-2020-038190

Hill-Briggs, F., Adler, N. E., Berkowitz, S. A., Chin, M. H., Gary-Webb, T. L., Navas-Acien, A., et al. (2021). Social determinants of health and diabetes: a scientific review. *Diabetes Care* 44, 258–279. doi: 10.2337/dci20-0053

Holt-Lunstad, J., Robles, T. F., and Sbarra, D. A. (2017). Advancing social connection as a public health priority in the United States. *Am. Psychol.* 72, 517–530. doi: 10.1037/amp0000103

Howick, J., Moscrop, A., Mebius, A., Fanshawe, T. R., Lewith, G., Bishop, F. L., et al. (2018). Effects of empathic and positive communication in healthcare consultations: a systematic review and meta-analysis. *J. R. Soc. Med.* 111, 240–252. doi: 10.1177/0141076818769477

Ikeda, A., Kawachi, I., Iso, H., Iwasaki, M., Inoue, M., and Tsugane, S. (2013). Social support and cancer incidence and mortality: the JPHC study cohort II. *Cancer Causes Control* 24, 847–860. doi: 10.1007/s10552-013-0147-7

Joshanloo, M., Jose, P. E., and Kielpikowski, M. (2017). The value of exploratory structural equation modeling in identifying factor overlap in the mental health continuum-short form (MHC-SF): a study with a New Zealand sample. *J. Happiness Stud.* 18, 1061–1074. doi: 10.1007/s10902-016-9767-4

Katz, M. R., Irish, J. C., Devins, G. M., Rodin, G. M., and Gullane, P. J. (2003). Psychosocial adjustment in head and neck cancer: the impact of disfigurement, gender and social support. *Head Neck* 25, 103–112. doi: 10.1002/hed.10174

Kawa, M. H. (2017). Influence of perceived social support and meaning in life on fighting spirit: a study of cancer patients. *Int. J. Adv. Educ. Res.* 2, 86–93.

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling.* Guilford Press.

Korkmaz, S., Goksuluk, D., and Zararsiz, G. (2019). MVN: an R package for assessing multivariate normality. *R J.* 6:151. doi: 10.32614/rj-2014-031

Kuuppelomäki, M. (2003). Emotional support for dying patients: the nurses' perspective. *Eur. J. Oncol. Nurs.* 7, 120–129. doi: 10.1016/S1462-3889(03)00002-4

Langford, C. P. H., Bowsher, J., Maloney, J. P., and Lillis, P. P. (1997). Social support: a conceptual analysis. *J. Adv. Nurs.* 25, 95–100. doi: 10.1046/j.1365-2648.1997.1997025095.x

Lauriola, M., and Tomai, M. (2019). Biopsychosocial correlates of adjustment to cancer during chemotherapy: the key role of health-related quality of life. *Sci. World J.* 2019:9750940. doi: 10.1155/2019/9750940

Magasi, S., Ryan, G., Revicki, D., Lenderking, W., Hays, R. D., Brod, M., et al. (2012). Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. *Qual. Life Res.* 21, 739–746. doi: 10.1007/s11136-011-9990-8

Makoul, G., Krupat, E., and Chang, C. H. (2007). Measuring patient views of physician communication skills: development and testing of the communication assessment tool. *Patient Educ. Couns.* 67, 333–342. doi: 10.1016/j.pec.2007.05.005

Malley, J., and Fernández, J.-L. (2010). Measuring quality in social care services: theory and practice. *Ann. Public Coop. Econ.* 81, 559–582. doi: 10.1111/j.1467-8292.2010.00422.x

Marsh, H. W., Morin, A. J. S., Parker, P. D., and Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annu. Rev. Clin. Psychol.* 10, 85–110. doi: 10.1146/annurev-clinpsy-032813-153700

Mazzoni, D., and Cicognani, E. (2016). Positive and problematic support, stress and quality of life in patients with systemic lupus erythematosus. *Anxiety Stress Coping* 29, 542–551. doi: 10.1080/10615806.2015.1134785

Mazzoni, D., Cicognani, E., and Prati, G. (2017). Health-related quality of life in systemic lupus erythematosus: a longitudinal study on the impact of problematic support and self-efficacy. *Lupus* 26, 125–131. doi: 10.1177/0961203316646459

Morin, A. J. S., Arens, A. K., and Marsh, H. W. (2016). A Bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Struct. Equ. Model. Multidiscip. J.* 23, 116–139. doi: 10.1080/10705511.2014.961800

Morin, A. J. S., Katrin Arens, A., and Marsh, H. W. (2015). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Struct. Equ. Model.* 23, 116–139. doi: 10.1080/10705511.2014.961800

Müller, E., Zill, J. M., Dirmaier, J., Härter, M., and Scholl, I. (2014). Assessment of trust in physician: a systematic review of measures. *PLoS One* 9:e106844. doi: 10.1371/journal.pone.0106844

Noble, L. M. (2020). "Doctor-patient communication and adherence to treatment," in *Adherence to Treatment in Medical Conditions*, CRC Press. 51–82. doi: 10.1201/9781003072348

Nouman, H., and Zanbar, L. (2020). Support or stressor? The community as a predictor of perceptions of infertility. *Soc. Work Health Care* 59, 650–667. doi: 10.1080/00981389.2020.1852360

Pinquart, M., and Duberstein, P. (2010). Associations of social networks with cancer mortality: a meta- analysis. *Crit. Rev. Oncol. Hematol.* 75, 122–137. doi: 10.1016/j.critrevonc.2009.06.003

Ranjan, C., and Najari, V. (2022). nlcor: Compute Nonlinear Correlations. R package version 2.3. [Source code] Retrieved from https://rdrr.io/github/ProcessMiner/nlcor/src/R/correlations.R

Reinikka, R., and Svensson, J. (2010). Working for god? Evidence from a change in financing of nonprofit health care providers in Uganda. *J. Eur. Econ. Assoc.* 8, 1159–1178. doi: 10.1111/j.1542-4774.2010.tb00551.x

Reynolds, J. S., and Perrin, N. A. (2004). Mismatches in social support and psychology adjustment to breast cancer. *Health Psychol.* 23, 425–430. doi: 10.1037/0278-6133.23.4.425

Rhemtulla, M., Brosseau-Liard, P. É., and Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* 17, 354–373. doi: 10.1037/a0029315

Rizalar, S., Ozbas, A., Akyolcu, N., and Gungor, B. (2014). Effect of perceived social support on psychosocial adjustment of Turkish patients with breast cancer. *Asian Pac. J. Cancer Prev.* 15, 3429–3434. doi: 10.7314/APJCP.2014.15.8.3429

Rodriguez, A., Reise, S. P., and Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *J. Pers. Assess.* 98, 223–237. doi: 10.1080/00223891.2015.1089249

Rutten, L. J. F., Arora, N. K., Bakos, A. D., Aziz, N., and Rowland, J. (2005). Information needs and sources of information among cancer patients: a systematic review of research (1980–2003). *Patient Educ. Couns.* 57, 250–261. doi: 10.1016/j.pec.2004.06.006

Scala, D., Menditto, E., Armellino, M. F., Manguso, F., Monetti, V. M., Orlando, V., et al. (2016). Italian translation and cultural adaptation of the communication assessment tool in an outpatient surgical clinic. *BMC Health Serv. Res.* 16:163. doi: 10.1186/s12913-016-1411-9

Schetter, C. D. (2017). Moving research on health and close relationships forward-a challenge and an obligation: introduction to the special issue. *Am. Psychol.* 72, 511–516. doi: 10.1037/amp0000158

Sebri, V., Mazzoni, D., Triberti, S., and Pravettoni, G. (2021). The impact of unsupportive social support on the injured self in breast cancer patients. *Front. Psychol.* 12:722211. doi: 10.3389/fpsyg.2021.722211

Seemann, A.-K., Drevs, F., Gebele, C., and Tscheulin, D. K. (2015). Are religiously affiliated hospitals more than just nonprofits? A study on stereotypical patient perceptions and preferences. *J. Relig. Health* 54, 1027–1039. doi: 10.1007/s10943-014-9880-9

Sijtsma, K., and Molenaar, I. (2002). *Introduction to Nonparametric Item Response Theory*. Thousand Oaks: Sage Publications.

Straat, J. H., van der Ark, L. A., and Sijtsma, K. (2016). Using conditional association to identify locally independent item sets. *Methodology* 12, 117–123. doi: 10.1027/1614-2241/a000115

Street, R. L., Makoul, G., Arora, N. K., and Epstein, R. M. (2009). How does communication heal? Pathways linking clinician-patient communication to health outcomes. *Patient Educ. Couns.* 74, 295–301. doi: 10.1016/j.pec.2008.11.015

Strom, J. L., and Egede, L. E. (2012). The impact of social support on outcomes in adult patients with type 2 diabetes: a systematic review. *Curr. Diab. Rep.* 12, 769–781. doi: 10.1007/s11892-012-0317-0

Suurmeijer, T. P. B. M., Doeglas, D. M., Briançon, S., Krijnen, W. P., Krol, B., Sanderman, R., et al. (1995). The measurement of social support in the

'European research on incapacitating diseases and social support': the development of the social support questionnaire for transactions (SSQT). *Soc. Sci. Med.* 40, 1221–1229. doi: 10.1016/0277-9536(94)00253-P

Thoits, P. A. (2011). Mechanisms linking social ties and support to physical and mental health. *J. Health Soc. Behav.* 52, 145–161. doi: 10.1177/0022146510395592

Tomai, M., Lauriola, M., and Caputo, A. (2019). Are social support and coping styles differently associated with adjustment to cancer in early and advanced stages? *Mediterr. J. Clin. Psychol.* 7, 71–24. doi: 10.6092/2282-1619/2019.7.1983

Usta, Y. Y. (2012). Importance of social support in cancer patients. *Asian Pac. J. Cancer Prev.* 13, 3569–3572. doi: 10.7314/APJCP.2012.13.8.3569

Van Dam, H. A., Van Der Horst, F. G., Knoops, L., Ryckman, R. M., Crebolder, H. F. J. M., and Van Den Borne, B. H. W. (2005). Social support in diabetes: a systematic review of controlled intervention studies. *Patient Educ. Couns.* 59, 1–12. doi: 10.1016/j.pec.2004.11.001

van der Ark, L. A. (2007). Mokken scale analysis in R. *J. Stat. Softw.* 20, 1–19. doi: 10.18637/jss.v020.i11

Venkatesh, S., and Weatherspoon, L. (2013). Social and health care provider support in diabetes self-management. *Am. J. Health Behav.* 37, 112–121. doi: 10.5993/AJHB.37.1.13

Wang, X., Cai, L., Qian, J., and Peng, J. (2014). Social support moderates stress effects on depression. *Int. J. Ment. Heal. Syst.* 8:41. doi: 10.1186/1752-4458-8-41

Wardian, J., and Sun, F. (2014). Factors associated with diabetes-related distress: implications for diabetes self-management. *Soc. Work Health Care* 53, 364–381. doi: 10.1080/00981389.2014.884038

Wenrich, M. D., Curtis, J. R., Ambrozy, D. A., Carline, J. D., Shannon, S. E., and Ramsey, P. G. (2003). Dying patients' need for emotional support and personalized care from physicians: perspectives of patients with terminal illness, families, and health care providers. *J. Pain Symptom Manag.* 25, 236–246. doi: 10.1016/S0885-3924(02)00694-2

White, K. R., and Begun, J. W. (1998). How does catholic hospital sponsorship affect services provided? *Inquiry* 35, 398–407. PMID: 10047770

White, K. R., Chou, T. H., and Dandi, R. (2010). Catholic hospital services for vulnerable populations: are system values sufficient determinants? *Health Care Manag. Rev.* 35, 175–186. doi: 10.1097/HMR.0b013e3181cafa20

Whitley, R. (2012). Religious competence as cultural competence. *Transcult. Psychiatry* 49, 245–260. doi: 10.1177/1363461512439088

Yağmur, Y., and Duman, M. (2016). The relationship between the social support level perceived by patients with gynecologic cancer and mental adjustment to cancer. *Int. J. Gynecol. Obstet.* 134, 208–211. doi: 10.1016/j.ijgo.2015.12.010

Zhou, P., Bundorf, K., Le Chang, J., Huang, J. X., and Xue, D. (2011). Organizational culture and its relationship with hospital performance in public hospitals in China. *Health Serv. Res.* 46, 2139–2160. doi: 10.1111/j.1475-6773.2011.01336.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership