



OPEN ACCESS

EDITED BY

Elizabeth Archer,
University of the Western Cape,
South Africa

REVIEWED BY

Juliette Lyons-Thomas,
Educational Testing Service,
United States
Yi-Hsin Chen,
University of South Florida,
United States

*CORRESPONDENCE

Karolina Lamtiur Dalimunthe
✉ karolina@unpad.ac.id

SPECIALTY SECTION

This article was submitted to
Assessment, Testing and Applied Measurement,
a section of the journal
Frontiers in Education

RECEIVED 26 December 2022

ACCEPTED 28 February 2023

PUBLISHED 17 March 2023

CITATION

Susanto H, Wedyaswari M and
Dalimunthe KL (2023) A content validity and
cognitive interview to develop the HARMONI
items: Instrument measuring student well-
being in West Java, Indonesia.
Front. Educ. 8:1132031.
doi: 10.3389/feduc.2023.1132031

COPYRIGHT

© 2023 Susanto, Wedyaswari and Dalimunthe.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

A content validity and cognitive interview to develop the HARMONI items: Instrument measuring student well-being in West Java, Indonesia

Hery Susanto, Miryam Wedyaswari and
Karolina Lamtiur Dalimunthe*

Center of Psychological Innovation and Research, Faculty of Psychology, Universitas Padjadjaran,
Sumedang, Indonesia

Practitioners in education need instrument to measure the skills of well-being. These skills should be taught in school alongside achievement skills to improve the learning process and student outcome. However, most existing well-being instruments are designed for the western population, even though geographic regions and cultural diversities are important to consider to develop well-being instruments. Thus, we use the PROSPER (Positivity, Relationships, Outcomes, Strengths, Purpose, Engagement, Resilience) model as a student well-being framework to develop an instrument named the HARMONI (*Hasil yang berproses, Andal berdaya lenting, Relasi yang positif, Makna dalam tujuan, Orientasi sikap positif, Nilai suatu kekuatan, Inisiatif yang melibatkan*), which is specified for West Java high school student population. The HARMONI instrument has seven dimensions: Positivity, Relationships, Outcomes, Strengths, Purpose, Engagement and Resilience. The study aims to develop HARMONI items for high school students in West Java. We conducted three activities: item development, expert reviews, and cognitive interviews. In the review process, nine experts in total who are experts in educational psychology, well-being, and psychometry were involved. Based on the reviews, we reduced the items from 556 to 125 items. The calculation of the Content Validity Index of the reviews showed that the items have good content validity (mean I-CVI ranging from 0.96 to 1). Next, cognitive interviews were conducted to analyze participants' cognitive processes in responding to the items. The participants of the cognitive interviews were 25 senior high school students in West Java, who were selected based on a purposive sampling technique. The result showed that the participants well understood most of the items. However, several items needed to be revised due to problems related to comprehension, retrieval and judgment. The items were improved by changing incomprehensible, vague and ambiguous words, revising the sentence structure and adding more specific cues. A total of 124 items were ready to be used for a psychometric study.

KEYWORDS

student well-being, cognitive interview, content validity, PROSPER framework, Indonesia

1. Introduction

Student well-being is an important aspect of students' lives. Thus, the "skills of well-being" are needed to teach alongside achievement skills. Well-being should be taught in school to prevent mental health problems, provide a better learning process and increase life satisfaction (Seligman et al., 2009). Promoting well-being in school has become the focus of studies and interventions in many countries through the positive education approach (e.g., Bonell et al., 2016; Coulombe et al., 2020; Murray and Fordham, 2020).

In Indonesia, the promotion of student well-being has also gained attention in educational settings. In 2018, the regional government of West Java launched an education program called Jabar Masagi, which aimed to strengthen the foundation of the young generation in West Java through character education and was targeted to influence the social behaviors of high school students. The program embedded local wisdom by including West Java values as the basis of the student well-being concept (Jabar, 2021). In developing the curriculum, the program employed the PROSPER framework as a reference for curriculum development. The PROSPER is an acronym for a multidimensional concept of well-being, including Positive emotions, Relationships, Outcomes, Strengths, Purpose, Engagement, and Resilience. The PROSPER was chosen because it provides the school pathways to improve the well-being of the students through the integration of hedonic and eudaimonism views of well-being (Noble and McGrath, 2015). In Jabar Masagi, the term was translated into HARMONI in Bahasa Indonesia, which also reflects the same ideas about student well-being as thriving and succeeding healthily as an intended purpose of the well-being framework.

Although the promotion of well-being has been embedded in the education program in West Java high schools, no instrument has been developed to measure the well-being of the students based on the PROSPER framework. Most existing well-being instruments are intended for the western population, and when applied to the non-western population, it has some limitations due to geographic regions and cultural diversities. The two aspects are important to develop well-being instruments (Lindert et al., 2015). A previous study suggests that variability in well-being scores across countries may be attributed to cultural differences. For example, Asian students may report scores closer to neutral thus lowering the overall scores of well-being. One possible explanation for such a problem is that individuals from collectivist cultures are not supposed to be distinguished (Diener et al., 1995). The impact of culture on the conception and measurement of well-being has also been discussed, including the definition of happiness, response style, memory, and judgmental biases (Oishi et al., 2013; Oishi, 2018). In Indonesia, students' behavior is directed toward the attainment of personal and collective well-being. The impact of culture on well-being measures can be identified in how students set their purpose in life, express themselves, and interact. In our previous study, we found that students set their purposes according to internalized cultural values and norms that they think will ultimately make their parents happy. In terms of expression, students feel more grateful than proud of their accomplishments. Some students consider feeling proud an inappropriate expression because it could convey an arrogant attitude that is against local values. Lastly, students should apply tata krama (i.e., code of conduct and beliefs about moral values) to develop good relationships with teachers (Dalimunthe et al., 2022). To our

knowledge, no published research on developing the PROSPER instrument is specified for the Indonesian population. Thus, there remain challenges in how the PROSPER construct should be operationalized and measured by considering the characteristics of West Java students.

Furthermore, there have been some student well-being measurements besides PROSPER, such as PERMA (Positive emotions, Engagement, Relationships, Meaning and Purpose and Accomplishment) and SSWQ (Student Subjective well-being Questionnaire) translated and implemented as student well-being measurements in Indonesia. However, compared to those measurements, PROSPER provides a comprehensive framework in terms of two added components that cannot be found in other measurements: strength and resilience. With these two components, not only has PROSPER include character and ability strength in its framework, but it also considers the student's capability to withstand adversities, thus allowing them to feel good and function well. With the last two additional components, PROSPER can overcome other measurement limitations that do not consider the student's capabilities to bounce back in a difficult situation using their strengths which is an important component of well-being (Noble and McGrath, 2015).

For instrument development, we define student well-being as "a relatively consistent mental and emotional condition characterized by positive feelings and attitudes, positive relationships with others in the school environment, resilience, optimal self-potential development, and a higher level of satisfaction with learning experience." (Dalimunthe et al., 2022). This definition contains two main components of well-being, which are (1) the fulfillment of satisfaction feelings in engaging in school activities, e.g., positive feeling and (2) the efforts to be better in school performance, e.g., applying character strengths on school activities. Similar to the PROSPER, the HARMONI student well-being instrument consists of seven dimensions called *Hasil yang berproses* (Outcomes), *Andal berdaya lenteng* (Resilience), *Relasi yang positif* (Relationships), *Makna dalam tujuan* (Purpose), *Orientasi sikap positif* (Positivity), *Nilai suatu kekuatan* (Strengths), and *Inisiatif yang melibatkan* (Engagement) (Noble and McGrath, 2015; Dalimunthe et al., 2022).

Each of the PROSPER dimensions is defined as follows. *Hasil yang berproses* (Outcomes) is a feeling of accomplishment and progress toward goals accompanied by various efforts and strategies that support progress and continuous self-development. *Andal berdaya lenteng* (Resilience) is one's capacity to "bounce back" when they face a challenging situation. *Relasi yang positif* (Relationships) is experiencing ongoing positive relationships with peers, teachers and school staff based on prosocial values. *Makna dalam tujuan* (Purpose) is the evaluation, perception and beliefs of students that the goals they set and the activities they conduct at school are meaningful and valuable for themselves and society. *Orientasi sikap positif* (Positivity) is a sustainable positive emotional state that results from applying a positive mindset when dealing with various school situations. *Nilai suatu kekuatan* (Strengths) is understanding and accepting one's character strengths and ability strengths and understanding how to apply these strengths in different contexts. *Inisiatif yang melibatkan* (Engagement) refers to conditions that reflect students' connection to various academic and non-academic activities (Noble and McGrath, 2015; Dalimunthe et al., 2022).

Previous studies showed that designing new instruments requires item development and evaluation (Boateng et al., 2018). The item

development starts with choosing the framework for item specification, writing the items, and then reviewing the items. To ascertain the quality of the items, test developer reviews for content quality, clarity and construct-irrelevant aspects of content (AERA et al., 2014). Meanwhile, among the methods used to develop items, cognitive interviewing is a method for gathering direct input from students on the item format and content (Irwin et al., 2009). The method has become an essential step to develop standardized measures and serves as a source of validity (Irwin et al., 2009; Ryan et al., 2012; Boeije and Willis, 2013). Both steps also provide two sources of validity evidence, namely validity evidence based on the content and response process (AERA et al., 2014). The current study will focus on the HARMONI item development of student well-being instruments using the PROSPER framework through the following process: (1) review of the content quality and construct relevance using expert review and (2) identify response processes of test-takers through the cognitive interview.

2. Materials and methods

There were three activities that we conducted for this study. The first was item development, the second was expert review, and the third was cognitive interview. The Universitas Padjadjaran Research Ethics Committee No.879/UN6.KEP/EC.2020 approved the study protocol.

2.1. Item development

Items were written in Bahasa Indonesia by nine individuals who have qualifications of bachelor's degrees in psychology and have participated in well-being research in the Indonesian educational context. To ensure the quality of the items, we held a workshop for the item writers. The workshop consisted of several activities, including discussions to understand the well-being construct; discussions to understand the definitions and examples of the PROSPER dimensions (i.e., Positivity, Relationships, Outcomes, Strengths, Purpose, Engagement, and Resilience); explanations about how to write the items; and exercises to write the items. We also specified that the items should represent the seven PROSPER dimensions and target West Java high school students. We employed a 5 points Likert-type scale from 1 = not at all like me to 5 = very much like me for all the items. Each item writer wrote 5–10 items per dimension for 3 weeks. This activity produced 551 items from seven dimensions of PROSPER.

2.2. Expert review

We conducted the expert reviews in three rounds. The research team, which consists of three educational psychologists who are experts in student well-being construction and have a background in test construction, conducted the first review. Each member reviewed the items independently to assess the item relevance to measure each PROSPER construct. Item Content Validity Index (I-CVI) was computed for each item based on the rating of the item relevance. We employed a 4-point rating scale with 1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, and 4 = highly relevant. Then, for each

item, the I-CVI is computed as the number of reviewers giving a rating of either 3 or 4, which is then divided by the number of reviewers. If the I-CVI is lower than 1, the item is considered irrelevant to measure the PROSPER and is subjected to revision or elimination (Lynn, 1986). We also calculated the Mean I-CVI for each PROSPER dimension and the whole scale. Finn's coefficients using a two-way model were calculated as indices of inter-rater reliability (IRR) (Finn, 1970). Negative values indicate a poor IRR, values between 0.00 and 0.20 indicate a slight IRR; values between 0.21 and 0.40 indicate a fair IRR; values between 0.41 and 0.60 indicate a moderate IRR; values between 0.61 and 0.80 indicate a substantial IRR; and values between 0.81 and 1.00 indicate an almost perfect IRR (Landis and Koch, 1977).

In the second review, we invited three experts in student well-being, educational psychology, and psychometry to review the accepted items from the first round of review. The same review process and analysis procedure were employed in this round. The rating scale was provided to measure the relevancy of the construct, including the language use of each item. In addition, the experts provided recommendations concerning the revision of the items.

After the second round of review, the COVID-19 pandemic happened, and we had to adjust the content of the items so that they could be responded to by the students who learn from home. The adjustment was conducted by the research team so that the items can be used to measure student well-being, not only in a traditional learning situation (face-to-face learning situation) but also in a learning-from-home situation (e.g., "I have positive emotions at school" change to "I have positive emotions while studying"). After the adjustment was made, we conducted a third review by inviting three different experts using the procedure we used in the second round of review. This process produced 125 items.

2.3. Cognitive interview

A cognitive interview was conducted to analyze participants' cognitive process in responding to the items. We analyzed 125 items consisting of 15 items of Positivity (P), 19 items of Relationships (R), 19 items of Outcomes (O), 18 items of Strengths (S), 13 items of Purpose (PU), 23 items of Engagemet (E) and 18 items of Resilience (RE).

Twenty-five senior high school students in West Java were recruited for the interviews based on a purposive sampling technique. First, we selected the school area based on its representativeness of the West Java culture. Schools in Bandung, Bekasi and Cirebon represent the Parahyangan, Betawi and Cirebon cultures, respectively. The selection of the students was conducted through the recommendation from school officials based on the representativeness of gender, grade, school type (i.e., general or vocational high school) and school area (i.e., rural or urban).

The students who participated in the interviews were native Indonesian speakers who could read Indonesian texts. Participants were required to have sufficient facilities such as laptops or other gadgets, good internet networks and the ability to operate the technology because interviews were conducted online. If the participants did not have these facilities, investigators contacted their schools to facilitate the students.

The individual interviews were designed to explore students' cognitive processes as they understand the items, recall experiences

needed to respond to the item, decide which choice is more relevant to them, and select an appropriate and meaningful response (Tourangeau et al., 2004).

The reparative approach was used in the cognitive interview. This approach focused on improving the questionnaire items and reducing errors in responses. The cognitive interview used verbal probing techniques through a retrospective procedure in which the interviewers asked the participants to fill in all items on the questionnaire first and then asked questions that explored the participant's cognitive process. Individuals who have experience in conducting psychological assessments and have received training on using the cognitive interview protocol conducted the cognitive interview one-on-one. The entire process was conducted online using the Google Meet application at the location that the participant specified.

The interview was conducted in two parts. In the first part, the interviewer introduces him/herself and then explains the informed consent contents. It included the interview purpose, the benefits of participating, the reward, the risks that might be experienced, the recording of the interview process, and the confidentiality of the data. If the participant agreed to participate, the interviewer guided them to fill in personal information data and the HARMONI items. Each participant filled in two dimensions of the PROSPER framework of student well-being. The investigators ensured that seven to eight participants responded to each item. During this part, participants turned on audio, and video, and shared the questionnaire display screen. All the processes were recorded. The first part of the interview took about 20 min.

In the second part, the interviewer performed verbal probing for a maximum of 2 h (see Table 1). During this part, video and audio were activated to capture significant and possibly relevant verbal and nonverbal expressions. The recording features in Google Meet were used to record the interviews. Then, the interviews were transcribed verbatim for analysis.

A top-down approach with cognitive coding techniques was employed to analyze the interview data. Analysis was performed on each item in the aggregate. Coding was made on each item based on the conclusion of the interview transcript per item from all participants. The cognitive coding scheme was based on Tourangeau (1984) four steps of cognitive coding, which include: (1)

comprehension of the question, (2) retrieval of relevant information, (3) judgment process, and (4) response process. With this cognitive coding, we could find the source(s) of the problem in each item that is subjected to revisions (Willis, 2015).

3. Results

In this section, we present the study's results including the analyzes of the expert reviews and the cognitive interviews.

3.1. Expert reviews

Table 2 shows the mean I-CVI for each item for each round of expert review. The first review had Mean I-CVI for the seven dimensions ranging between 0.47 and 0.58, which indicated that revisions of the items were needed. Finn's coefficients ranged between 0.41 and 0.57, which indicated moderate agreement among experts. Four hundred thirty-two items were dropped during this round due to the low quality of the items including the language of the items, relevancy of the items with the PROSPER construct and redundancy of the items. Thus, 124 items were kept. The second round of expert review had Mean I-CVI ranging from 0.72 to 0.98, which indicated that, in general, the items were relevant for measuring each PROSPER dimension. Finn's coefficients ranged between 0.33 and 0.82, which indicated a fair to almost perfect agreement among experts. In the third round, Mean I-CVI ranged between 0.96 and 1, which indicated that, overall, the quality of the revised items improved. Experts in this round showed better agreement on the relevance of the items as indicated by Finn's coefficients that range between 0.69 and 0.84 (substantial to almost perfect agreement). The notes from the experts provided suggestions related to the relevancy of the item content to measure each well-being dimension. Additionally, suggestions were also given to revise the language of the items, including the use of unfamiliar words and ineffective sentence structure. In the final round of review, the following revisions were made to the 10 items with I-CVI values of less than 1, which are P2, P3, R13 R16, O1, O2, S16, E18, RE8, and RE16 (see Supplementary Table S1). We also decided to add one item in the Outcomes dimension to check for the best

TABLE 1 Cognitive interview guide based on Tourangeau Cognitive model (Willis, 2015).

Cognitive operations	Objective(s)	Question(s)
Understanding/Comprehension	To check participants' perception of the item.	What is the meaning of this item? What do you think this item talks about? Can you explain this item to me?
	To know what specific words and phrases in the item mean.	Is there any difficult word in this item? Inquiry: what does that word mean?
Recall/Retrieval	To identify what types of information the participant needs to respond to the item.	What event and experience was remembered when you read this item? Inquiry: tell me more about that event/experience.
Judgment	To know how much the participant devotes sufficient mental effort to respond to the item thoughtfully.	How difficult is it for you to decide the answer? Please explain the reasons.
	To know how confident the participant is in deciding the response.	How confident are you with your choice? Please explain the reason.
Response	To know how the participant chooses his/her responses and distinguish between choices.	Why do you choose this response for this item? Inquiry: how do you distinguish between 1/2/3/4/5 responses in this item?

TABLE 2 Number of items, mean item-content validity index (I-CVI), and Finn's coefficients.

Dimension	First review			Second review			Third review		
	Item (n)	Mean I-CVI	Finn ₂	Item (n)	Mean I-CVI	Finn ₂	Item (n)	Mean I-CVI	Finn ₂
Positivity	128	0.55	0.57	13	0.95	0.63	15	0.96	0.73
Relationships	75	0.47	0.50	20	0.88	0.55	19	0.96	0.78
Outcomes	80	0.52	0.51	18	0.85	0.54	18	0.96	0.73
Strengths	77	0.52	0.41	19	0.72	0.33	18	0.98	0.78
Purpose	63	0.58	0.53	12	0.92	0.62	13	1.00	0.84
Engagement	64	0.54	0.52	24	0.93	0.62	23	0.99	0.77
Resilience	64	0.48	0.48	18	0.98	0.82	18	0.96	0.69
TOTAL	556	0.53	0.51	124	0.89	0.58	124	0.97	0.76

I-CVI, Item content validity index; n, number of items; Finn₂, Finn's coefficient with a two-way model.

translation of “pride” in Bahasa Indonesia as suggested in the reviews. The items obtained from the expert reviews were evaluated in terms of the validity of the response process.

3.2. Cognitive interview

The results of the cognitive interview analysis are presented based on the cognitive coding that describes each cognitive operation in the response process, namely (1) comprehension of the question, (2) retrieval of relevant information, (3) judgment process, and (4) response process. The summary of the coding results for the 125 items is presented in Table 3. A total of 16 items were identified as problematic with four of the items having more than one source of the problem (printed in bold), which are items R19, O7, S1, and RE12. The wording of these items is provided in Supplementary Table S2. To provide a comparison with the expert review result, we also added information for items with low I-CVI (printed in underline). Two items with low I-CVI were still identified as problematic in the cognitive interview process (items O2 and E18). The following paragraphs present the results from each cognitive coding. In each section, examples will be presented in English with the original wording given in square brackets.

3.2.1. Comprehension

The comprehension stage is the first stage in the cognitive process when participants respond to the items. At this stage, they had to understand the intent of the statement before they could recall the experience that the item was expected to bring. The results of coding on cognitive interview data showed that some of the items in the HARMONI compiled are quite difficult for participants to understand. Several sources of problems that cause difficulty in comprehension were found.

The first problem was the items that contained words that were difficult for participants to understand. An example of an item with this problem was item P5 (“I enjoy participating in learning activities that the teacher designed”). Two participants stated that they were confused about the meaning of the word “designed” [dirancang]. One participant suggested replacing it with the word “planned” [direncanakan/ditentukan] to gain an understanding of the item while the other participant suggested replacing it with the word “arranged”

TABLE 3 Coding summary.

Cognitive coding	Item
Comprehension of the question (CQ)	P5, P10, R19 , <u>O2</u> , O12, S13, S15, <u>E18</u> , RE12
Retrieval of relevant information (RI)	R19 , O7 , O16, O17, S1 , S2, S3, RE12
Judgment process (JP)	O7 , O8, S1
Response process (RP)	–

P, prosper; R, relationships; O, outcomes; S, strengths; PU, purpose; E, engagement; RE, resilience; bold, item with more than one source of problem; underline, item with low I-CVI.

[disusun]. Another example of an item with this problem was item E18 (“I took the initiative to invite friends to do activities together”), which contained the word “initiative” [inisiatif]. One participant said that the word “initiative” was difficult to understand while several others interpreted the word as desires [keinginan], ideas [pemikiran baru/ide] and thoughts [nalar]. Other words that some of the participants did not understand are the words “obstacles” [hambatan] (P10), “think” [beranggapan] (P10), “capacity” [kapasitas] (S3), “area” [area] (S5), “context” [konteks] (S15), “initiative” [inisiatif] (E18), “expectation” [harapan] (O12), “alternative” [alternatif] (RE12), “shame” [malu] (S13), “learning activities” [kegiatan belajar] (P5), “success” [keberhasilan] (O2), “a lot” [banyak] (O2) and “complaints” [komplain] (R19).

The second problem was related to syntax complexity. This problem was related to the difficulty of interpreting one part of the sentence, for example, “I think there is a lesson that I can take in every obstacle I meet during school” in item P10. In this item, a participant said that the sentence was too convoluted. Other sentences that were difficult to understand were “...maintaining the spirit achieving hope during school” [mempertahankan semangat dalam mencapai harapan saya selama bersekolah] (O12), “... to achieve expectation” [mencapai harapan] (O16), “...ways to help” [cara untuk membantu] (O17), “... character strength” [kekuatan dari karakter] (S1), and “... rarely feel ashamed” [jarang merasa malu] (S13).

3.2.2. Retrieval

The problem at this stage occurs when the statement presented to the participants fails in triggering the long-term memory to retrieve

relevant experiences. The previous process also caused some of the problems in this process. The problem of retrieving experience could also occur when the cue presented on the item was not sufficient and when the participant had never had the experience relevant to the item. An example of an item with this source of the problem was item R19 (“School employees listen to student complaints”). All participants said that they never experienced complaining to school staff. Another example was item O7 (“I am disciplined conduct the plans that I have drawn up”). Some participants had difficulty responding to the item because they never made a plan. A similar problem was also found in item O16 (“I can use certain skills to achieve my expectations during school”). In this item, some participants could not recall memories of skills. In addition, the problem to retrieve also occurred in other items: item O17 (“I apply many ways to help me achieve my targets during school”), S1 (“I recognize my character strength”), S2 (“I understand myself”), S3 (“I know my capacity to solve problems well”), RE12 (“I have alternative ways to overcome the obstacles I experience”).

3.2.3. Judgment process

Problems at this stage occur when participants have difficulty in making decisions and evaluating their responses to the items. The item that indicated a problem in this process was item O7 (“I am disciplined conduct the plans that I have drawn up”). On this item, some participants found it difficult to estimate and determine their level of discipline. Thus, the difficulty made them confused when assigning the rating. Some participants tried to calculate whether they made plans and how many times they conducted the plans. Another item that also showed this source of the problem was O8 (“I study very hard to get good results”). Almost like the previous item, on this item, some participants found it difficult to determine how hard “very hard” was. A participant said that it took a while for him to decide because he thought that good outcomes did not have to accompany studying optimally. Participants found it difficult to determine the response because they were confused when choosing whether to use the learning outcomes or the efforts they make as a reference to determine the response. Problems in the judgment process were also found in item S1 (“I recognize my character strength”). In this item, there was one participant who admitted that it was difficult for him to decide because he did not know himself yet.

3.2.4. Response process

Problems related to the response process occurred when the participants found it difficult to choose the scale after making some considerations. In the present study, the processes at the previous stages, especially at the judgment stage mainly caused the difficulties in choosing a response on a scale. In the cognitive interviews, no specific problems were identified as being caused by the scale provided. All participants were able to explain the basis for their response mappings and to distinguish between each response category very well. The answers to the interview questions also showed that the direction of the responses is per the estimation made by the participants. The following responses on item R5 provide an example of such answers:

[(Question: “Why did you choose response two?”) ... because to me response two describes my situation well. I cannot tell everything to my friends. Some things are secret, some things can

be shared with my friends. But it is still possible for me to share some stories. I chose response two because I regard myself as an introverted person. (Inquiry: “Why didn’t you choose response one?”) I didn’t choose response one because there are moments when I share stories just to motivate my friends, mostly my juniors.”].

[(Question: “What was your reason for choosing response four between one and five?”) “... Umm, because response five means I share too much with my friends; like I tell everything. Well, I don’t say everything. So, some things I keep for myself. I didn’t choose response five because it means I tell them everything. I don’t say everything.”].

4. Discussion

This study aims to develop the HARMONI items through content validity and cognitive interview. In the present study, the developed items were considered relevant to measure each dimension of the PROSPER. Based on the results, we found that expert review identified problems related to content relevancy and semantic aspects of the items. Based on the cognitive interview, we also found that high school students in West Java were able to respond to most items measuring each PROSPER dimension. This finding is consistent with other studies targeted at adolescent populations. Previous studies show that adolescents tend to better understand items compared to younger children (e.g., [Irwin et al., 2009](#); [Ravens-Sieberer et al., 2014](#)). The high school student participants well comprehended most of the items. However, we also identified several terms that the students poorly understood, thus inhibiting their ability to recall any relevant experience and proceed to the following cognitive process. Our result is consistent with findings from previous studies that cognitive interviews are more likely to detect problems related to comprehension ([Presser and Blair, 1994](#); [Willis et al., 1999](#)). We found two major problems categorized as comprehension problems: word familiarity and sentence complexity. In our study, the student participants seemed confused about the terms having vague meanings or words seldom used in everyday situations. They also had difficulty understanding an item using a compound sentence. Comprehension problems were identified in almost every PROSPER dimension.

In addition to comprehension problems, our study also identified problems related to the retrieval process. The student participants experienced retrieval problems when the retrieval cue provided in the presented items was not enough to recall experiences stored in long-term memory ([Tourangeau et al., 2004](#)). In our study, several items were identified as providing irrelevant cues to the participants. Therefore, they could not respond to the items. Additionally, several items were quite vague concerning the cues presented. Failure in the participants’ recalling process may happen when incongruity occurs between the cues and their personal experience or when the cues do not help them identify the relevant experience ([Tourangeau, 1999](#)). We found that some of the items intended to measure the Relationships, Outcomes, Strengths and Resilience dimensions of well-being are not relevant to high school students’ learning

experience because most of them never had the experience or did not recognize the situation provided.

The final problem identified in the cognitive interview is related to the judgment process. The participants estimate the extent to which the description in each statement applied to them. We found that the difficulty in making decisions on the response happened when the participants felt unsure about the reference for their estimation, especially in estimating certain qualities, such as discipline, hardness, strengths and character, which seemed abstract and immeasurable. Some participants tried to calculate the frequency of their behaviors to perform the estimation. A previous study assumed that participants tried to recall the occurrences of behavior when they were asked to arrive at a frequency judgment (Menon and Yorkston, 1999). The failure to recall past behaviors may also have caused the problem found in this study as they did not have a clear reference. Problems in judgment were only identified in two dimensions, namely Outcomes and Strengths.

The current study identified four major problems in the HARMONI items covering content relevance, comprehension, recall and judgment cognitive problems. Each problem may have a different impact on the measurement process. First, it is important for a new instrument that individual responses are aligned with the instrument's purpose to describe appropriate psychological processes and to direct accurate intervention or policy for the targeted population in the future (Beauchamp and McEwan, 2017). Secondly, the poorly comprehended items will lead to errors in choosing the relevant experiences and errors in responses to the items (Willis, 2015). Thirdly, problems in the retrieval process will determine the relevancy of experience retrieved from long-term memory (Jobe et al., 1993; Tourangeau et al., 2004). Lastly, the error in the judgment process will lead to a wrong evaluation of oneself, which raises doubts about whether the score describes the individual accurately (Zumbo and Hubley, 2017). Remedies should be made to the items to enhance their quality.

Based on the identified problems, specifically through cognitive interviews, we provided several remedies for the problematic items. A previous study shows that small changes in the appearance of words can lead to greater clarity in understanding items (Zamanzadeh et al., 2022). We made several revisions to the current study. Firstly, the comprehension problems were revised by replacing the complicated or vague words with more familiar words and changing the problematic sentence structure. Then, retrieval cues that are more relevant to students' experiences were employed to revise items with retrieval problems. Finally, we use more specific and measurable cues to replace abstract concepts (i.e., adjectives) to enhance each participant's capability to judge their responses. 16 items were revised in the process. We also decided to drop item R19 because all the participants could not respond to the item. We concluded that the revision to this item would not improve the quality of the items. This process resulted in the final 124 items being ready for large-scale testing.

In line with previous literature, we found that expert reviews and cognitive interviews provide a different source of information for item revisions. Expert reviews provided information about the relevancy of the content based on the experts' ratings. In addition, cognitive interviews provided information on the cognitive process of the participants during the survey. Previous studies showed that different methods for item development tend to show

low consistencies since they provide a different source of problems (Presser and Blair, 1994; Rothgeb et al., 2005; Yan et al., 2012). However, we found that combining the two methods is very useful in the item development process. Expert reviews identified problems related to content relevancy and semantics in the very early stage of the item development. Meanwhile, the application of cognitive coding as a follow-up step helped identify more specific problems according to the complex cognitive response process.

The study has several implications. To begin with, this study demonstrates how content validity and cognitive interview can be used in the early stage of the development process of an instrument. Then, this study provides items to measure well-being based on the PROSPER framework in Bahasa Indonesia that may be used for further development processes (i.e., psychometric testing). Moreover, this study provides two sources of validity evidence for the HARMONI instrument, namely evidence-based on the content and evidence-based on the response process. Both sources of evidence will support the use of the instrument to measure student well-being in Indonesia. Finally, this study provides information on the potential problems in item development and possible related remedies. This information can become a valuable resource for other researchers who develop instruments measuring student well-being, especially in the Indonesian context.

Two limitations were identified in the current study. Firstly, the data in this study were collected from a sample drawn from a limited number of areas in West Java and thus may affect the results. Students from different areas may have different characteristics and show variations in their response process. Secondly, seven to eight participants responded to each item in this study. Although this number may be sufficient, some authors recommend using 12–20 participants to respond to each item (Guest et al., 2006; Willis, 2015). Finally, each participant in the cognitive interview only responded to two dimensions of the PROSPER. Therefore, some important findings to evaluate the whole well-being construct could be missed. Further studies can be conducted by collecting data from more representative areas of West Java to evaluate the psychometric properties of the items and thus provide more information on the quality of the items. Such studies should also provide other sources of validity evidence that the Standard recommends (AERA et al., 2014).

Overall, the findings of the content validity and cognitive interviews suggest that high school students in West Java were able to respond to most items measuring each PROSPER dimension (see [Supplementary Table S3](#) for good items example). The final 124 items in this study are ready for large-scale testing to evaluate the instrument's psychometrics.

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 Universitas Padjadjaran Research Ethics Committee. The

patients/participants provided their written informed consent to participate in this study.

Author contributions

HS, MW, and KD contributed to the study conception, design, writing of the manuscript, and data analysis. KD and HS arranged the the concepts and design. MW and HS conducted the literature search. Enumerators who KD and HS trained did data acquisition. MW did manuscript preparation, HS did manuscript editing, and KD did manuscript review. All authors contributed to the article and approved the submitted version.

Funding

The author(s) disclosed receipt of the following financial support for the research of this article: the Education Office of West Java Province included in the Project “Jabar Masagi” supported this research.

Acknowledgments

The authors would like to acknowledge the contributions of Sudarmo Wiyono, M.Si., Psikolog., Surya Cahyadi, M.Psi., Psikolog.,

Theresia Novi Poespita Chandra, S.Psi., M.Si., Ph.D., Psikolog., Margaretha, S.Psi., PGDip., Whisnu Yudianta, M.Psi., Gr. Cert. Ed., Psikolog., Puspita Adhi Kusuma Wijayanti, M.Psi., Psikolog for their contributions to reviewing items.

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2023.1132031/full#supplementary-material>

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association. Washington, D.C.
- Beauchamp, M. R., and McEwan, D. (2017). “Response processes and measurement validity in health psychology” in *Understanding and Investigating Response Processes in Validation Research* (Cham: Springer), 13–30.
- 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, 1–18. doi: 10.3389/fpubh.2018.00149
- Boeije, H., and Willis, G. (2013). The cognitive interviewing reporting framework (CIRF): towards the harmonization of cognitive testing reports. *Methodology* 9, 87–95. doi: 10.1027/1614-2241/a000075
- Bonell, C., Hinds, K., Dickson, K., Thomas, J., Fletcher, A., Murphy, S., et al. (2016). What is positive youth development and how might it reduce substance use and violence? A systematic review and synthesis of theoretical literature health behavior, health promotion and society. *BMC Public Health* 16:135. doi: 10.1186/s12889-016-2817-3
- Coulombe, S., Hardy, K., and Goldfarb, R. (2020). Promoting well-being through positive education: a critical review and proposed social-ecological approach. *Theory Res. Educ.* 18, 295–321. doi: 10.1177/1477878520988432
- Dalimunthe, K. L., Susanto, H., and Wedyaswari, M. (2022). A qualitative study exploring the construct of student well-being in West Java high school students. *Psychol. Res. Urban Soc.* 5:1. doi: 10.7454/proust.v5i2.155
- Diener, E., Suh, E. M., Smith, H., and Shao, L. (1995). National differences in reported subjective well-being: why do they occur? *Soc. Indic. Res.* 34, 7–32. doi: 10.1007/BF01078966
- Finn, R. H. (1970). A note on estimating the reliability of categorical data. *Educ. Psychol. Meas.* 30, 71–76. doi: 10.1177/001316447003000106
- Guest, G., Bunce, A., and Johnson, L. (2006). How many interviews are enough?: An experiment with data saturation and variability. *Field Methods* 18, 59–82. doi: 10.1177/1525822X05279903
- Irwin, D. E., Varni, J. W., Yeatts, K., and DeWalt, D. A. (2009). Cognitive interviewing methodology in the development of a pediatric item bank: A patient reported outcomes measurement information system (PROMIS) study. *Health Qual. Life Outcomes* 7, 1–10. doi: 10.1186/1477-7525-7-3
- Jabar, A. H. (2021). *Cetak Generasi Muda Berkarakter Dengan Jabar masagi*. Available at: <http://humas.jabarprov.go.id/cetak-generasi-muda-berkarakter-dengan-jabar-masagi/2368#:~:text=Jabar%20Masagi%20merupakan%20program%20yang,kearifan%20lokal%20jabar%20menjadi%20dasarnya> (Accessed October 26, 2021).
- Jobe, J. B., Tourangeau, R., and Smith, A. F. (1993). Contributions of survey research to the understanding of memory. *Appl. Cogn. Psychol.* 7, 567–584. doi: 10.1002/acp.2350070703
- Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310
- Lindert, J., Bain, P. A., Kubzansky, L. D., and Stein, C. (2015). Well-being measurement and the WHO health policy health 2010: systematic review of measurement scales. *Eur. J. Pub. Health* 25, 731–740. doi: 10.1093/eurpub/cku193
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nurs. Res.* 35, 382–385. doi: 10.1097/00006199-198611000-00017
- Menon, G., and Yorkston, E. A. (1999). “The use of memory and contextual cues in the formation of behavioral frequency judgments” in *The science of Self-report*. eds. A. A. Stone, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman and V. S. Cain (United Kingdom: Psychology Press), 75–92.
- Murray, L., and Fordham, L. A. (2020). *Promoting children's social and emotional wellbeing: evaluation of the Positive Living Skills Primary School Wellbeing Program in a New South Wales regional school*. Positive Living Skills.
- Noble, T., and McGrath, H. (2015). PROSPER: A new framework for positive education. *Psychol. Well-Being* 5, 1–17. doi: 10.1186/s13612-015-0030-2
- Oishi, S. (2018). “Culture and subjective well-being: conceptual and measurement issues” in *Handbook of Well-Being*. eds. E. Diener, S. Oishi and L. Tay (Salt Lake City, UT: DEF Publishers)
- Oishi, S., Graham, J., Kesebir, S., and Galinha, I. C. (2013). Concepts of happiness across time and cultures. *Personal. Soc. Psychol. Bull.* 39, 559–577. doi: 10.1177/0146167213480042
- Presser, S., and Blair, J. (1994). Survey pretesting: do different methods produce different results? *Sociol. Methodol.* 24:73. doi: 10.2307/270979
- Ravens-Sieberer, U., Devine, J., Bevans, K., Riley, A. W., Moon, J., Salsman, J. M., et al. (2014). Subjective well-being measures for children were developed within the PROMIS project: presentation of first results. *J. Clin. Epidemiol.* 67, 207–218. doi: 10.1016/j.jclinepi.2013.08.018

- Rothgeb, J., Willis, G., and Forsyth, B. (2005). Questionnaire pretesting methods: do different techniques and different organizations produce similar results? *96 Annual Conference of Americ.* 31–35.
- Ryan, K., Gannon-Slater, N., and Culbertson, M. J. (2012). Improving survey methods with cognitive interviews in small-and medium-scale evaluations. *Am. J. Eval.* 33, 414–430. doi: 10.1177/1098214012441499
- Seligman, M. E. P., Ernst, R. M., Gillham, J., Reivich, K., and Linkins, M. (2009). Positive education: positive psychology and classroom interventions. *Oxf. Rev. Educ.* 35, 293–311. doi: 10.1080/03054980902934563
- Tourangeau, R. (1984). “Cognitive science and survey methods” in *Cognitive aspects of survey methodology: Building a bridge between disciplines*. eds. T. Jabine, M. Straf, J. Xanur, and R. Tourangeau (Washington, DC: National Academy Press), 73–100.
- Tourangeau, R. (1999). “Remembering what happened: memory errors and survey reports” in *The Science of Self-Report* (United Kingdom: Psychology Press), 41–60.
- Tourangeau, R., Rips, L. J., and Rasinski, K. (2004). *The Psychology of Survey Response*. Cambridge University Press. Cambridge, England.
- Willis, G. B. (2015) in *Analysis of the Cognitive Interview in Questionnaire Design*. ed. P. Leavy (England: Oxford University Press)
- Willis, G. B., Schechter, S., and Whitaker, K. (1999). A comparison of cognitive interviewing, expert review, and behavior coding: what do they tell us. *Am. Stat. Assoc. Proc. Section Surv. Res. Methods* 9, 28–37.
- Yan, T., Kreuter, F., and Tourangeau, R. (2012). Evaluating survey questions: a comparison of methods. *J. Off. Stat.* 28, 503–529.
- Zamanzadeh, V., Ghahramanian, A., and Valizadeh, L. (2022). Improving the Face Validity of Self-Report Scales through Cognitive Interviews Based on Tourangeau Question and Answer Framework: A Practical Work on the Nursing Talent Identification Scale. *Res. Square*. doi: 10.21203/rs.3.rs-587444/v1
- Zumbo, B. D., and Hubley, A. M. (2017). *Understanding and investigating response processes in validation research*. In Springer. New York City. (Vol. 69, Issue).