



# Further Insights Into the Beck Hopelessness Scale (BHS): Unidimensionality Among Psychiatric Inpatients

Michela Balsamo<sup>1\*</sup>, Leonardo Carlucci<sup>1</sup>, Marco Innamorati<sup>2</sup>, David Lester<sup>3</sup> and Maurizio Pompili<sup>4</sup>

<sup>1</sup> Department of Psychology, Health and Territory, School of Medicine and Health Sciences, "G. d'Annunzio" University of Chieti-Pescara, Chieti, Italy, <sup>2</sup> Dipartimento di Scienze Umane, Università Europea di Roma, Rome, Italy, <sup>3</sup> Stockton University, Galloway, NJ, United States, <sup>4</sup> Department of Neurosciences, Mental Health and Sensory Organs, Suicide Prevention Center, Sant'Andrea Hospital, Sapienza University of Rome, Rome, Italy

## OPEN ACCESS

### Edited by:

Xenia Gonda,  
Semmelweis University, Hungary

### Reviewed by:

Roger Watson,  
University of Hull, United Kingdom  
Adriana Mihai,  
University of Medicine and Pharmacy  
of Târgu Mureș, Romania

### \*Correspondence:

Michela Balsamo  
mbalsamo@unich.it

### Specialty section:

This article was submitted to  
Public Mental Health,  
a section of the journal  
Frontiers in Psychiatry

**Received:** 10 January 2020

**Accepted:** 09 July 2020

**Published:** 31 July 2020

### Citation:

Balsamo M, Carlucci L, Innamorati M, Lester D and Pompili M (2020) Further Insights Into the Beck Hopelessness Scale (BHS): Unidimensionality Among Psychiatric Inpatients. *Front. Psychiatry* 11:727. doi: 10.3389/fpsy.2020.00727

Short versions of the Beck Hopelessness Scale have all been created according the Classical Test Theory, but the use and the application of this theory has been repeatedly criticized. In the current study, the Item Response Theory approach was employed to refine and shorten the BHS in order to build a reasonably coherent unidimensional scale whose items/symptoms can be treated as ordinal indicators of the theoretical concept of hopelessness, scaled along a single continuum. In a sample of 492 psychiatrically hospitalized, adult patients (51.2% females), predominantly with a diagnosis of Bipolar Disorder type II, the BHS was submitted to Mokken Scale Analysis. A final set of the nine best-fitting items satisfied the assumptions of local independency, monotonicity, and invariance of the item ordering. Using the ROC curve method, the IRT-based 9-item BHS showed good discriminant validity in categorizing psychiatric inpatients with high/medium suicidal risk and patients with and without suicide attempts. With high sensitivity (>.90), this newly developed scale could be used as a valid screening tool for suicidal risk assessment in psychiatric inpatients.

**Keywords:** hopelessness, inpatients, depression, Mokken analysis, unidimensionality

## INTRODUCTION

Hopelessness is an important psychological construct, defined as negative expectations regarding oneself and one's future life and a negative emotional state characterized by the lack of finding a solution for one's problems (1). In his research focused on depression mood and suicidal behavior, Beck (2–5) observed that patients diagnosed with depressive disorders shared common cognitive features—a negative view of the self, and of the self in relation to the world and in relation to the future. He paid special attention to one of these cognitive features—a negative view of the self in relation to the future, by introducing the construct of “hopelessness”. Unlike depression, it is oriented towards the future as opposed to the present state (6).

According to Beck and his associates, hopelessness has substantial clinical utility for suicide risk assessment and prediction of future suicidal behavior. They produced empirical evidence for the

association between hopelessness and suicidality by arguing that severity of suicidal intent is more strongly related to hopelessness compared to depression (5, 7–11).

Subsequently, research showed that hopelessness could lead to suicidality (12–14). As a modifiable key psychological risk factor in suicidal behavior, with an impact that can be reduced by means of appropriate psychotherapeutic interventions, the recognition and assessment of hopelessness plays an important role in the prevention of suicidal behavior (15–17).

## Measurement of Hopelessness: The Beck Hopelessness Scale

To investigate better the construct of the hopelessness, Beck (4), Beck, Weissman, et al. (6) constructed the Beck Hopelessness Scale (BHS). In its development, Beck grouped 9 items from an unpublished inventory assessing the attitudes about the future and 11 items drawn from a set of pessimistic statements formulated by patients with psychiatric diagnoses (6). The BHS scores were found to be strongly correlated with clinical ratings of hopelessness by Beck and colleagues in their validation study (6).

Several studies have indicated good predictive validity for the BHS (11, 14, 18–20). For instance, the BHS was found to predict suicidal thoughts and attempts among 289 psychiatrically hospitalized suicidal youth across a 1–6-month follow-up after hospital discharge (21, 22). Hopelessness, as measured by the BHS, was found a significant predictor of attempted suicides among psychotic patients at first admission to hospital (18).

The BHS performed similarly across inpatients and outpatients, for both psychiatric and medical samples (23–25), and can also be used for predicting social functioning and general status health in psychiatric samples (26).

## Overview on Dimensionality BHS

In order to determine the dimensionality of the scale, Beck, Weissman, et al. (6) subjected the items of the BHS to a factor analysis. The factors were labeled “Feelings about the Future”, “Loss of Motivation”, and “Future Expectations”, respectively. According to the authors, although the factor structure of the BHS made sense clinically, it can vary according to the type of clinical sample being studied and the type of factor-analytic methods conducted. Further studies analyzed this factor structure across different samples (27–30). According to the review by Aish et al. (31), factor structures found in the literature could be grouped as follows: (1) one-factor models (32); (2) two-factor models (33–38); (3) three factor models (27, 29, 30); and (4) models with four or more factors (28). In the reported studies, the emerging factors found differed from those identified by Beck et al.’s study in terms of the assigned factors’ labels and their item composition.

In addition, some authors noted that some original items could fit models different from those proposed (31, 32, 36). For example, Aish and Wasserman (31) found that no strong evidence supported the multidimensionality of the BHS by the first CFA. In detail, 15 items tapped a single dimension of hopelessness, and so a reduced number of 4 items could summarize most of the information contained by the BHS. Thus, the dimensionality of the BHS remains an open issue of

considerable interest. For example, Hill, Gallagher, et al. (27) found that only one component (giving up - the motivational component) was significantly related to suicidal intent. In this case, combining different dimensions into a composite scale (39) might reduce the predictive validity of the BHS.

## Short Versions of the BHS

The length of the BHS could be discouraging for the respondents. Lengthy questionnaires reduce data quality and respondent willingness (21), especially in clinical populations (40, 41). In order to be useful in practical settings, an instrument should be sufficiently brief and easy to complete (42, 43), especially when multiple measurement scales are employed.

Previous methodological studies have suggested that a reduction by about 50–70% of the number of items could not compromise substantially the original psychometric functioning of a scale (44–47). This is also true for the BHS (31, 48). For example, some researchers have suggested that a single item, “My future seems dark to me” (item #7), could be sufficient to assess hopelessness. According to Aish and Wasserman (31), this sentence is ideal for summarizing the construct under investigation: the perception of a menacingly ambiguous and hopeless future. This suggestion was supported by Perczel Forintos et al.’s (13) study whose results showed that this item had the highest item-residue correlation ( $r = .75$ ), that is the highest correlation with the total score of the BHS.

Other researchers have investigated the psychometric properties of different 3- or 4-item versions of the BHS. Based on confirmatory factor analysis, Aish and Wasserman (31) reported excellent fit for a 4-item version of the BHS (composed of items #6, 7, 9, 15). In a cross-sectional survey, Yip and Cheung (1) administered this shortened version to some 2,000 Chinese subjects. A significantly high correlation ( $r = .88$ ,  $p < .001$ ) of the shortened version with the original 20-item BHS was found, suggesting that the abbreviated scale can be reliably used in clinical studies. They also reported that the 4-item BHS was able to differentiate patients with and without suicidality, similarly to the original version of the BHS. Recently, Aloba, Akinsulore (48) introduced a new 4-item version of the BHS (composed of items 8, 9, 13, and 15) in a sample of 327 Nigerian adult psychiatric outpatients. The authors reported satisfactory reliability and validity, comparable to that of the long form of the BHS. Other researchers (28, 37) have also suggested that a three-item version of the BHS (items 7, 14, and 20) could represent the scale and be a valid measure of hopelessness.

Finally, some researchers have devised brief modified versions of the BHS. For example, Perczel Forintos, Sallai (13) proposed a three-item version of the BHS in their study on a clinical sample of 300 individuals. Three items with highest correlations with the BHS total score, plus an item # 2 from the Beck Depression Inventory (BDI), which refers to hopelessness, were included in this brief BHS. Scores on this scale were highly correlated with scores on the original scale ( $r = .88$ ) and had relatively high internal consistency (Cronbach’s  $\alpha$  coefficient:  $r = .80$ ). More recently, Fraser, Burnell, et al. (49) developed two short hopelessness measures by re-wording two items of the BHS

negatively (Brief-H-Neg) and items positively worded (Brief-H-Pos), and shifting the response format from “yes/no” to 5-point Likert scale (from “absolutely agree” to “absolutely disagree”). Nevertheless, no strong methodological evidence (i.e., construct validity) can be found in the literature for the Brief-H-Pos/Neg short forms. In addition, these two short forms could potentially be affected by the reverse-item bias, which is very common in scale with Likert response format (50).

## Aims

Short versions of the BHS have all been created according the Classical Test Theory (CTT) (31, 48, 49), despite the fact that the use of CTT has been criticized (51–53). The Item Response Theory (IRT) approach to the refinement of measures of clinical constructs has many practical advantages (54). For example, IRT methods could: (i) detect subtle changes in patients’ mental health that would not be recorded with the use of the mean or summed scores; (ii) overcome the sample dependence found in CTT; and (iii) produce invariant item/person statistics that allow optimal individual scores and comparison of individual scores across different tests (55, 56). In addition, applying item response models to the validation of psychopathology measures can help build a “*reasonably coherent unidimensional scale*” [(57), p.475] and treat symptoms as ordinal indicator of risk scaled along a single continuum. The unidimensional assumption was rarely met (58), especially using the CTT framework. The BHS, in this context, is not an exception. Thus, applying IRT models to the BHS could improve Beck’s conceptualization of hopelessness as a unidimensional measure.

Therefore, the purposes of the current study were: (1) to investigate the psychometric properties of the individual items of the BHS; (2) to develop a reliable and valid version including a reduced set of items since time-effective instruments would be of great practical value both in clinical and research settings; and (3) to test the diagnostic performance of the proposed short version of the BHS in classifying psychiatric inpatients at higher risk of suicide, and to compare its performance with the original 20-item BHS and to other short versions proposed in the literature (31, 48). The versions proposed by Fraser, Burnell, et al. (49) were excluded in the comparison analysis due to their methodological weaknesses and the lack of studies in support of their psychometric validity.

## METHODS

### Participants and Procedure

The sample included 492 psychiatrically hospitalized, adult patients of whom 48.8% were males, with a mean age of 39.09 (SD = 13.13) years. Participants were recruited from January 2014 to December 2018 at psychiatric units situated in Sant’Andrea Medical Center, an affiliate of “Sapienza” University of Rome, Italy.

Inclusion criteria were to be inpatients aged 18 years or over with current psychiatric diagnosis performed according to the criteria of Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR). Diagnoses were made by expert clinicians within the first 48 h of the psychiatric

hospitalization. These were supported by means of an examination according to the Mini International Neuropsychiatric Interview criteria (59) and administered a psychological battery of tests, including the BHS, to assess the severity of psychopathology, and the presence of risk factors for suicide. The information was retrieved from clinical files for the indicated period of time. Participants with cognitive impairment and degenerative neurological disease were excluded from the study.

Primary psychiatric diagnoses included 174 (35.4%) patients with Bipolar Disorder type I (BD-I), 52 (10.6%) patients with diagnosis of Bipolar Disorder type II (BD-II), 58 (11.8%) patients with Major Depressive Disorder (MDD), 83 (16.9%) patients with Psychosis, 66 (13.4%) patients with schizoaffective disorder, 41 (8.3%) with other Axis I disorders, and 18 (3.7%) with no DSM-IV-TR diagnosis. In the total sample, 21 (4.3%) had a Personality Disorder (PD) as a secondary DSM-IV-TR psychiatric diagnosis.

The patients participated voluntarily and provided written informed consent. The study protocol was reviewed and approved by the local research ethics review board, with assurance that data would be reported anonymously and in aggregate form. All procedures were in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments.

## Measure

### Beck Hopelessness Scale (BHS)

The BHS is composed of 20 dichotomous “true/false” items that aimed to assess three major aspects of hopelessness: feelings about the future, loss of motivation, and expectations. Total scores were created by first reverse-coding nine items (items 1, 3, 5, 6, 8, 10, 13, 15, 19) and then summing the item scores. Higher total scores indicate greater hopelessness (range 0–20). The Italian version of the BHS has been translated and validated by Innamorati, Lester, et al. (60) with the permission of Pearson Education (Upper Saddle River, NJ 07458, USA). A series of studies has shown that the BHS performed similarly across psychiatric inpatients and outpatients and medical samples (23–25, 36).

## Data Analysis

The Mokken Scale Analysis (MSA) was carried out within the framework of Non-parametric Item Response Theory (NIRT), in order to (a) evaluate the fundamental measurement properties of the BHS; (b) address dimensionality issue problems raised from previous research; and (c) refine the scale by providing a unidimensional, brief and reliable measure. Compared to the parametric IRT models, the Mokken probabilistic approach does not require strict assumptions about the data, and persons are allocated to a finite number of discrete ability levels. Thus, the relationship between the latent variables and the probability for a response were not required to match a specific shape (61). For this reason, Mokken’s model has been considered as less parsimonious than Rasch model. However, as pointed out by Emons, Sijtsma and Meijer (62) and Wind (63), the application of parametric IRT models (the Rasch model) might lead to inappropriate conclusions in: a) diagnosing psychological

latent variables that are not clearly understood, and b) assessing the monotonicity assumption when it does not hold for a particular item (63). In this view, Mokken non-parametric models represents a viable alternative to Rasch model (64).

The MSA for dichotomous response items includes the evaluation of two models: the Monotonic Homogeneity Model (MH) and the Double Monotonicity Model (DM) (61, 65–67). Briefly, the MH model entails an ordinal scale person measurement (68), which means that the relative ordering of psychiatric inpatients on the hopelessness latent variable is invariant across items. Data are found to fit the MH model if three underlying assumptions are satisfied: *Monotonicity*, *Unidimensionality*, and *Local independence*.

The DM model represents a special case of the MH model. In our case, assessing the DM model means that all the BHS items were ordered in the same way across the psychiatric inpatients. In addition to the MH assumptions, a fourth assumption is required for the DM model: the *Invariant Item Order* (IIO). Since the DM model provides evidence for invariant ordering of items and sample for dichotomous items, this model best represents the ordinal version of the Rasch model or the 1PL-IRT (69, 70).

Like other IRT models (e.g., the Rasch model), the MSA involves an iterative process in which an observed pattern of responses is refined in order to reach the overall fit to the model expectations.

Following Sijtsma, Meijer (71) and Sijtsma and Molenaar (65), a series of steps were carried out in order to assess both the MH and the DM models as well as the scale properties (i.e., the reliability). All analyses were performed using the *Mokken* package of R (72, 73).

- The *Unidimensionality* assumption was assessed by performing the Automated Item Selection Procedure (AISP) algorithm using consecutively different values of  $c$  (.30, .35, .40, .45, .50) (74). Once a scale was detected, the scalability coefficients for individual items ( $H_i$ ), item pairs ( $H_{ij}$ ), and for the total scale ( $H$ ) were computed, along with the standard error. For  $H$  and  $H_i$ , values of  $H \geq 3$  identify a “sufficient” scalability. For  $H_{ij}$ , values greater than 0 and positive identify good scalability (61).
- *Local independence* was investigated using the Straat, van der Ark (75) conditional association procedure and the  $W_1$  and  $W_3$  indices developed in the *Mokken* package. The procedure identifies locally dependent item pairs.
- The *Monotonicity* of the Item Response Function (IRF) was assessed using a non-parametric regression method (76). To evaluate if IRFs were non-decreasing and monotonic, we took into account the size of the violation of monotonicity ( $\#vi/\#ac$ ) which should not exceed the value of .3; and the Diagnostic Crit Value (crit) which should not exceed the value of 80.
- The *Invariant Item Ordering* (IIO) assumption was investigated using the method MIIO (Manifest IIO) (77). Violations of the IIO were assessed by taking into account the size of “ $\#vi/\#ac$ ” and the “Crit” indexes. When the IIO has been established, the coefficient  $H^T$  expresses the precision of

the item ordering (from 0 “weak” to 1 “high precision”, with a minimum value of .3).

- *Reliability* was assessed using the Moolenaar–Sijtsma method (MS) (78). Cronbach (79) alpha and the Latent Class Reliability Coefficient (LCRC) (80) were also computed.
- Next, we compared the resulting unidimensional shortened version of the BHS with the Hungarian 4-item BHS models developed by Aish and Wasserman (31) and Aloba, Akinsulore (48), in order to assess which of the three competitive brief versions of the BHS performs better in measuring Beck’s Hopelessness.

The diagnostic performance of the refined 9-item BHS was assessed using the Area Under (AUC) the receiver operating characteristic curve (ROC). The Youden (J) method was employed in order to detect the cut-off score of the final item set, and we also computed key predictive statistics, including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). ROC curve analysis was done using the MedCalc software package (MedCalc software, Mariakerke, Belgium) (81).

Optimal values of AUC ranged from 0 “weak performance” to 1 “perfect performance” (82), with a recommended value of  $>.70$  (83). The MINI Suicidal Subscale (59) cut-off score was employed to classify participants with high and moderate suicidal risk. We also computed a series of pairwise comparison of ROC curves to test whether the BHS long and short forms differed in performance across diagnoses. Finally, we tested the diagnostic accuracy of the 9-item BHS in discriminating between inpatients with and without prior suicide attempts.

## RESULTS

### Mokken Analysis

First, we re-coded all the BHS items written in a reversed format. Inadmissible scores as well as missing data were removed from the dataset. Next, we submitted the BHS items to a Mokken analysis to verify scalability and the unidimensionality assumption. As shown in **Table 1**, the individual item scalability ( $H_i$ ) of the 20-item BHS was below the accepted cut-off of 30 for items #1, 3, 5 and 13. Several  $H_{ij}$  coefficients were found to be negative and below the .03 cut-off (e.g., the paired items 1-3; 3-5; and 13-2). The  $H$  coefficient of .323 ( $\pm .02$ ) suggested a ‘weak’ scale and, therefore, was likely multidimensional. As expected, the AISP, with different values of lower bound, suggested a three-scale structure. The main scale was composed of 16 items identified as the “Hopelessness” dimension, while the remaining scales were small and composed of two items for each (scale 2: item 8 and 13; scale 3: items 1 and 3). However, the results confirmed the BHS as a unidimensional scale since the typical outcome pattern was confirmed (65, 84) and was observed using the AISP algorithm with different values of  $c$ . Hence, four items were discarded from the full 20-item scale and the remaining items were submitted to a MSA to explore the fit both of the MH

**TABLE 1 |** Descriptive statistics of the items and the scale for the 20-item-BHS, the refined 9-item and the Aish et al. (31) and Aloba et al. (48) 4-item models.

	20-item BHS model			refined 9-item model			Aish et al. (31) 4-item model			Aloba et al. (48) 4-item model		
	Scale	Hi	SE	Scale	Hi	SE	Scale	Hi	SE	Scale	Hi	SE
BHS2	1	.363	.025	1	.459	.032						
BHS4	1	.350	.046									
BHS5	1	.283	.030									
BHS6	1	.403	.038	1	.459	.052	1	.500	.056			
BHS7	1	.399	.023				1	.402	.035			
BHS9	1	.308	.027				2	.298	.041	1	.166	.041
BHS10	1	.304	.028									
BHS11	1	.393	.025	1	.489	.031						
BHS12	1	.397	.029	1	.499	.034						
BHS14	1	.389	.045	1	.505	.053						
BHS15	1	.349	.025				1	.363	.038	1	.327	.035
BHS16	1	.327	.027	1	.475	.035						
BHS17	1	.346	.026	1	.497	.033						
BHS18	1	.491	.028	1	.566	.036						
BHS19	1	.383	.031									
BHS20	1	.393	.026	1	.551	.032						
BHS8	2	.221	.029							1	.259	.036
BHS13	2	.291	.042							1	.402	.052
BHS1	3	.061	.032									
BHS3	3	.088	.038									
H (SE)		.323 (.017)			.500 (.024)			.375 (.034)			.275 (.033)	
H <sup>T</sup>		.37			.42			.34			.24	
MS		.87			.86			.64			.53	
$\alpha$		.86			.86			.60			.51	
LCRC		.89			.89			.63			.56	

*H<sub>i</sub>*, item scalability coefficient; *SE*, Standard Error of item scalability; *MS*, Molenaar–Sijtsma method;  $\alpha$ , Cronbach's alpha; *LCRC*, Latent Class Reliability Coefficient.

(unidimensionality, monotonicity, local independency) and the DM (invariant item ordering, IIO) Mokken's model, as well as to measure the reliability of the scale.

For the 16 item-BHS scale, the H-coefficient was .42 ( $\pm$  .02), and *H<sub>i</sub>* coefficients ranged from .323 (.034; item 5) to .554 (.029, item 18). All *H<sub>ij</sub>* coefficients were non-negative, but some paired items showed scalability coefficients below the threshold (*H<sub>ij</sub>* > .30). The conditional association procedure used to detect local independency suggested that the item pairs (4 with 7-11-12-14-18-19; 5-18; 6-18; and 14 with 14-18; 16-20; 19-18) were positively locally dependent. Next, the data analysis supported monotonicity, since no monotonicity violations were detected across all the items. Non-significant IIO was identified for items 12 and 6, and backward selection suggested removing item 7 (#vi/#ac = .20; crit = 73) and item 5 (#vi/#ac = .16; crit = 70), both of which showed signs of violating item ordering close to the recommended thresholds. The remaining items showed crit values <40, that indicated the violations reported were potentially due to sampling variations. The #vi/#ac values ranged from .9 to .2. H<sup>T</sup> was .373 indicating low accuracy of the item ordering. Reliability estimates were satisfactory, with an MS index of .87, a Cronbach  $\alpha$  of .86, and an LCRC of .89. Taken together, these results provided evidence for the weak unidimensionality of the Hopelessness scale, as it was composed at this stage. The MH and the DM Mokken's model requirements were partially met since neither local independency nor IIO was reached at this stage.

Next, we removed in turn the items labelled as locally dependent through the conditional association procedure and with the lower *H<sub>i</sub>* values, and the data set was iteratively

reanalyzed. Then we removed items that showed the greatest violation of item ordering.

### The Refined 9-Item Hopelessness Model

The refined Hopelessness scale resulted in a unidimensional set of nine items (items 2, 6, 11, 12, 14, 16, 17, 18, 20). The H-coefficient was .50 ( $\pm$  .02), all *H<sub>i</sub>* coefficients were greater than .46, and all *H<sub>ij</sub>* coefficients were non-negative. No violations of local independency and monotonicity were identified. A non-significant IIO was identified for all the items, and backward selection did not suggest removing any items. H<sup>T</sup> was .42 indicating medium accuracy of the item ordering. These results suggested that the refined Hopelessness scale was unidimensional and met the requirements of a MH and DM Mokken scale, although the scale's ability to discriminate between levels of hopelessness severity among psychiatric inpatients was medium. Concerning the scale properties, reliability estimates were satisfactory with an MS index of .86, a Cronbach  $\alpha$  of .86, and an LCRC of .89.

### Comparison of Brief Versions of the BHS

Finally, the Aish and Wasserman (31) and Aloba, Akinsulore (48) 4-item versions were submitted to the MSA in order to test which of the brief version best measured hopelessness. Concerning the Aish and Wasserman (31) model, the AISP algorithm suggested a two-dimensional scale structure. The main scale was composed of items 6, 7 and 15, while item 9 loaded on a separate dimension. The H-coefficient was .37 ( $\pm$  .03), all *H<sub>i</sub>* coefficients were greater than .23, and all *H<sub>ij</sub>* coefficients were

non-negative, with item 6 ( $SE = .05$ ) falling below the cut-off if the standard error was taken into account. No violations of monotonicity were identified, and no local dependence issue was observed. A significant violation of the IIO was identified for item 9, and backward selection suggested removing it from the model.  $H^T$  was .34 indicating poor accuracy of the item ordering. Reliability analysis suggested sufficient reliability for the scale, with an MS of .64, a Cronbach  $\alpha$  of .60, and an LCRC of .63. Taken together, the brief version of BHS proposed by Aish and Wasserman (31) satisfied the monotonicity and item local independency assumptions of Mokken analysis, but failed to address unidimensionality and the invariant item ordering, but displayed a sufficient reliability score.

With regard to the Aloba, Akinsulore (48) 4-item version, the H-coefficient was .27 ( $\pm .03$ ). Only two  $H_i$  coefficients were greater than .30, and all  $H_{ij}$  coefficients were non-negative, with all the items ( $SE$  ranged from .05 to .09) falling below the cut-off if the standard error was taken into account. No violations of monotonicity were identified, and a significant IIO was identified for items 9, 8 and 15.  $H^T$  was .24, indicating unacceptable accuracy of the item ordering. In addition, a local dependence issue was observed between item 15 with items 9–13. The reliability analysis also suggested poor reliability for the scale, with an MS of .53, a Cronbach  $\alpha$  of .51, and an LCRC of .56. Thus, the Aloba, Akinsulore (48) model cannot be considered as a Mokken's reliable and valid measure of Beck's Hopelessness, and so we eliminated this scale from subsequently analyses.

In conclusion, the refined 9-item model proposed here best represents a reliable and Mokken's suitable measure of Hopelessness compared to the Aish and Wasserman's (31) brief version.

## ROC Curve Analysis

A first ROC curve analysis was performed to compare the psychiatric inpatients with a high risk of suicide versus the low

risk group. The results indicated that the 9-item BHS scale was able to discriminate between the two groups. The AUC for the 9-item BHS total score was .708 (95%CI = .665–.748), suggesting good discrimination between the groups. The Youden index of .39 for the 9-item BHS total score was observed at a score of 3 points, corresponding to a sensitivity of 68.56% and specificity of 64.43%. Positive and negative predictive power were 55.6% and 75.9%, respectively.

Similarly, a second ROC curve was performed to compare the psychiatric inpatients with a medium risk of suicide versus the low risk group. The results indicated that the 9-item BHS scale was able to discriminate the two groups with an AUC of .522 (95%CI of .477–.567). The Youden index of .13 for the 9-item BHS total score was observed at a score of 1 point, corresponding to a sensitivity of 90.91% and specificity of 22.22%. The positive and negative predictive powers were 7.8% and 97.1%, respectively. Thus, better accuracy was displayed by the 9-item BHS brief version in correctly diagnosing psychiatric inpatients at high risk of suicide compared to those with a medium risk of suicide.

When we compared the predictive validity of the total scores of the BHS long form and the brief 9 and 4-item models, the AUCs were identical, and the differences among them were not found to be significant for those with a high risk of suicide ( $\Delta AUC$  ranged from .003 to .019) or for those with a medium risk of suicide ( $\Delta AUC$  ranged from .008 to .046). Thus, results of the pairwise comparison revealed that proposed 9-item BHS brief version did not differ in diagnostic accuracy from the 20-item long form or the Aish et al.'s 4-item short form. Indicators of the predictive accuracy of the BHS scales are shown in **Table 2**.

Concerning the ability of the 9-item BHS into discriminate between subgroups of inpatients with or without suicide attempts, no differences were found in the AUC. The results indicated that the 9-item BHS brief version was able to detect with the same accuracy psychiatric inpatients with versus without any previous suicide attempt.

**TABLE 2 |** Area Under the Curve (AUC) of the Receiver Operating Characteristic Curve (ROC) Analyses for the different version of the BHS, and Comparison of independent ROC curves.

MINI Suicidality Subscale – cut-off	N	AUC	SE	(95%CI)	Sensitivity/ Specificity	Cut-off	PPV/NPV	Model	$\Delta AUC$	SE	(95%CI)	z	p
High Risk	194												
20-item BHS		.724	.023	.682–.763	59.79/72.82	>8	58.9/73.6	BHS20 vs BHS9	.016	.011	–.005–.038	1.440	.150
9-item BHS		.708	.024	.665–.748	68.56/64.43	>3	55.6/75.9	BHS20 vs BHS4	.019	.013	–.006–.045	1.461	.144
4-item BHS		.705	.023	.662–.745	59.28/69.46	>1	55.8/72.4	BHS9 vs BHS4	.003	.017	–.032–.038	.181	.856
9-item BHS					N+	N-							
Inpatients with no attempt	55	.775	.078	.645–.875	11	46		with vs no attempt	.055	.135	–.209–.320	.409	.683
Inpatients with attempt	32	.720	.110	.527–.867	21	9							
Moderate Risk	33												
20-item BHS		.560	.047	.515–.604	81.82/32.90	>4	8.1/96.2	BHS20 vs BHS9	.037	.022	–.006–.081	1.674	.094
9-item BHS		.522	.047	.477–.567	90.91/22.22	>1	7.8/97.1	BHS20 vs BHS4	.008	.031	–.054–.072	.276	.782
4-item BHS		.569	.047	.524–.613	57.58/59.26	>1	9.2/95.1	BHS9 vs BHS4	.046	.043	–.039–.031	1.066	.286
9-item BHS					N+	N-							
Inpatients with no attempt	78	.529	.123	.393–.663	6	51		with vs no attempt	.081	.231	–.370–.533	.354	.723
Inpatients with attempt	9	.611	.195	.417–.783	3	27							

## DISCUSSION

The BHS has been recognized as a powerful tool for predicting suicidal risk in patients diagnosed with depression mood (19, 85–87). The BHS has been used extensively under the assumption that it captures a single dimension, hopelessness, composed of three components: affective, motivational, and cognitive (6). However, international studies have reported that the BHS could consist of from one (31) to four dimensions (28). It should be noted that these differed for label and items composition.

In the current study, we submitted the BHS to a Mokken Scale Analysis (MSA) as a method to overcome the limitations of the CTT. The MSA allows the ordering of individuals on the basis of their raw scores and addresses the BHS unidimensional issue. Not surprisingly, we found that all the BHS 20 items did not tap a single unidimensional factor, but rather formed three dimensions.

The analysis did not support the original affective, motivational and cognitive model, or a clear single dimension of hopelessness. In line with a study by Aish and Wasserman (31), most of the item tapped a single dimension (Scale 1 with 16 items). Scale 2 contained items 8 (cognitive domain) and 13 (affective domain), while Scale 3 was contained items 1 (affective domain) and 3 (motivational domain). The H values were stable with respect to the scalability level for items in the Scale 1. Items that made up Scale 2 and 3 were found to be not scalable, suggesting that these scales were a weak indicator of hopelessness or that the item wording is poor. Likewise, local independency nor IIO assumptions were reached for several items at this step of the analysis. This implies that the individual's responses to BHS items were dependent on the individual's level of the latent trait being measured (88), as well as the ordering of the items according to its severity (or mean score) being different (not invariant) for individuals at different trait levels (89). Consistent with these results, both the MH and DM Mokken model assumptions have not been reached or met for the 20-item BHS version in the present sample of Italian psychiatric inpatients.

In order to obtain a conceptually clear measure of hopelessness, we removed items with low scalability, local dependency and a not invariant item ordering. After removing eleven items, the one-scale model maintained its psychometric viability. This was not at all obvious, given that removing items means obtaining less information for each individual and may impair construct validity and reliability (54, 62). The process resulted in a unidimensional set of nine items (items 2, 6, 11, 12, 14, 16, 17, 18, 20) and the assumptions of local independency and monotonicity for the MSA were satisfied, as well as the invariant item ordering feature. Six items corresponded to the motivational component of hopelessness, as conceived by Beck, Weissman, et al. (6), two items were drawn from the cognitive component and a single item from the affective component. All the items showed medium to high scalability coefficients, and the ability to discriminate between psychiatric inpatients with different levels of suicide risk.

Previously, some authors have proposed extremely short versions of the BHS [e.g., 2-item version by Fraser et al. (49) and 4-item versions by Aish et al. (31) and Aloba et al. (48)]. In our sample, Aish et al.'s (31) and Aloba et al.'s (48) competing models

revealed psychometric weaknesses. The IRT refined short version of the BHS proposed in our study could represent a good compromise between time costs, and measurement precision.

The development and the use of short forms of measures has encountered contradictory views in the literature on clinical assessment (90). Overall, reasons why scores on short measures are likely to have less predictive validity than scores on longer inventories concern the poor sampling of the relevant behaviors (*construct underrepresentation bias*), and the lack of interest among researchers to improve the methodology of short-form development (i.e. *random measurement error issue*) (91).

In this view, this refined 9-item BHS has been developed using a sound item-development procedure: NIRT models. As suggested by Smith, McCarthy, and Anderson (90), the IRT based approach to short-form construction can lead to a shorter assessment without all the methodological issues that are often evident within CTT.

Similar, longer scales were likely to have greater content validity and higher reliability scores. To date, the overlap of content validity and scale length makes it difficult to determine “*whether the improved criterion validity of longer scales is the result of the improved reliability of measurement or the result of greater content validity*” (p. 884) (91). For example, the use of 1-item and 2-item measures has been found to increase both the Type 1-2 error rates, while slightly longer scales were found to increase the validity of study findings (92).

Similar to the content of long versions, our 9-item BHS tapped all the aspects of hopelessness: affective (item #6), motivational (item # 2–11–12–17–20), and cognitive (item #14–18) ones. Not surprisingly, the motivational aspect of this construct had greater weight, with five items that assess giving up (i.e. “*deciding not to want anything*”). As reported above, Hill, Gallagher et al. (27) found that only this was significantly related to suicidal intent. Indeed, “*loss of motivation has been found to represent the clinical picture of giving up, unpleasantness and darkness, which is a reality in suicide attempters*” (p. 142) (93).

The IRT-refined short version of the BHS also had good discriminant validity in categorizing psychiatric inpatients with high or medium suicidal risk, and patients with and without suicide attempts. Indeed, differences in diagnostic accuracy among the original 20-item version of the BHS, the four-item versions, and the IRT-refined nine-items short BHS were not significant. For high suicidal risk, a cut-off value scores >3 (with scores >1 for a moderate risk) for the refined nine-item short version of the BHS seem suitable, and this suggests that all patients with a total score of 3 should be referred for further risk assessment and management. Our IRT-refined short BHS had high sensitivity (>.90) and could be used as a valid screening tool for medium risk of suicide assessment across psychiatric inpatients.

Our results should be considered in light of four limitations. First of all, this sample consisted of a heterogeneous sample of adult psychiatric inpatients with and without suicidal attempts. Results using a more homogeneous sample of patients with mood disorders could differ, and the diagnostic accuracy of our IRT-refined nine-item short BHS reported in our study might be specifically related to the present population used. Authors of

study Reise and Waller (94) showed that items selected by NIRT models were able to discriminate within a particular range of the latent dimension or in a specific population of interest. This means that if one is interested into monitoring hopelessness in a population of healthy people, it is appropriate to have items that discriminate in the low-to-average trait range.

Second, for some authors, the Mokken analysis represents an explorative approach to the development and validation of clinical scales (95, 96). For example, Meijer and Baneke (97) recommended using NIRT models to investigate the data structure and to understand how items were functioning before applying parametric IRT. Future studies may apply parametric IRT models (e.g., two parameter [2PL] and four parameter [4PL] models) to test if the behavior of the specific responses may assume a specific logistic curve. For example, applying a 4PL model to the BHS items could reveal that the probability that individuals with severe hopelessness trait manifest a specific symptom less than the 100%.

Third, social desirability or other distortions in test responses could affect self-report measures and consequently our results. Fourth, we did not investigate the predictive validity, or further important aspects of validity of this IRT-refined 9-item short BHS. Establishing predictive validity between the self-report or screening tool and a criterion measure becomes mandatory (98). Future studies are necessary to test the present refined measure with an already well-established measure, i.e. to predict suicidal ideation and/or attitude in medical inpatients and outpatients.

## CONCLUSIONS

Since its development, the Beck Hopelessness Scale was tested across different patient groups. Its shortening without a substantial loss of its predictive validity would be extremely useful in vulnerable patients, such as those to which it is addressed.

Nine best-fitting items of the Beck Hopelessness Scale satisfied the assumptions of local independency, monotonicity, and invariance of the item ordering when all the items were submitted to Mokken Scale Analysis in a large sample of adult psychiatric inpatients.

## REFERENCES

1. Yip PS, Cheung YB. Quick assessment of hopelessness: a cross-sectional study. *J Health Qual Life Outc* (2006) 4(1):13. doi: 10.1186/1477-7525-4-13
2. Beck AT. Thinking and depression: I. Idiosyncratic content and cognitive distortions. *Arch Gen Psychiatry* (1963) 9(4):324–33. doi: 10.1001/archpsyc.1963.01720160014002
3. Beck AT. *Depression*. Clinical, Experimental & Theoretical Aspects New York: Harper & Row (1967).
4. Beck AT. The development of depression: A cognitive model. In: Friedman R, Katz M, editors. *Psychology of Depression: Contemporary Theory and Research*. Washington, DC: Winston (1974). p. 3–27.
5. Beck AT, Kovacs M, Weissman A. Assessment of suicidal intention: the Scale for Suicide Ideation. *J Consult Clin Psychol* (1979) 47(2):343. doi: 10.1037/0022-006X.47.2.343

The IRT-based 9-item BHS showed good discriminant validity in categorizing psychiatric inpatients with high/medium suicidal risk and patients with and without suicide attempts, with high sensitivity (>.90). Thus, it could be used as a valid screening tool for suicidal risk assessment among psychiatric inpatients.

To our best knowledge, this study is the first focused on the application of the Item Response Theory approach to the refinement and shortening of the BHS. Previous short versions of this scale were developed within the Classical Test Theory. However, with the Item Response Theory it is possible to build a reasonably coherent unidimensional scale whose items/symptoms can be treated as ordinal indicators of the theoretical construct of hopelessness, scaled along a single continuum.

## DATA AVAILABILITY STATEMENT

The datasets analyzed in this article are not publicly available. Requests to access the datasets should be directed to MI, marco.innamorati@unier.it; MP, maurizio.pompili@uniroma1.it.

## ETHICS STATEMENT

The study protocol was reviewed and approved by the local research ethics review board (Sant'Andrea Medical Center, an affiliate of the Sapienza University of Rome, Italy). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

LC and MB designed the study and conducted the statistical analyses. LC, MI, MP, and MB interpreted the data. LC, MB, MI, and DL drafted the manuscript. MP and MI recruited the sample and collaborated in editing the final manuscript. All authors contributed to the article and approved the submitted version.

6. Beck AT, Weissman A, Lester D, Trexler L. The measurement of pessimism: the hopelessness scale. *J Consult Clin Psychol* (1974) 42(6):861–5. doi: 10.1037/h0037562
7. Minkoff K, Bergman E, Beck AT, Beck R. Hopelessness, depression, and attempted suicide. *Am J Psychiatry* (1973) 130(4):455–9. doi: 10.1176/ajp.130.4.455
8. O'Connor RC, Armitage CJ, Gray L. The role of clinical and social cognitive variables in parasuicide. *Br J Clin Psychol* (2006) 45(4):465–81. doi: 10.1348/014466505X82315
9. Pollock L, Williams J. Problem-solving in suicide attempters. *psychol Med* (2004) 34(1):163–7. doi: 10.1017/S0033291703008092
10. Beck AT, Kovacs M, Weissman A. Hopelessness and suicidal behavior: An overview. *JAMA* (1975) 234(11):1146–9. doi: 10.1001/jama.234.11.1146
11. Beck AT, Steer R, Kovacs M, Garrison B. Hopelessness and eventual suicide: a 10-year prospective study of patients hospitalized with suicidal ideation. *Am J Psychiatry* (1985) 142(5):559–63. doi: 10.1176/ajp.142.5.559



12. Kovacs M, Beck AT, Weissman A. Hopelessness: An indicator of suicidal risk. *Suicide Life-Threat Behav* (1975) 5(2):98–103. doi: 10.1111/j.1943-278X.1975.tb00317.x
13. Perczel Forintos D, Sallai J, Rózsa S. Reliability and validity of the Beck Hopeless Scale. *Psychiatr Hungar* (2001) 16:632–43.
14. Beck AT, Brown G, Berchick RJ, Stewart BL, Steer RA. Relationship between hopelessness and ultimate suicide: A replication with psychiatric outpatients. *Am J Psychiatry* (1990) 147(2):190–296. doi: 10.1176/ajp.147.2.190
15. Hawton K, Sutton L, Haw C, Sinclair J, Harriss L. Suicide and attempted suicide in bipolar disorder: a systematic review of risk factors. *J Clin Psychiatry* (2005) 66(6):693–704. doi: 10.4088/JCP.v66n0604
16. Williams M. *Cry of pain*. London: Penguin (1997).
17. Wasserman D, Wasserman C. Examples of how to develop suicide prevention on all the continents. In: Wasserman D, Wasserman C, editors. *Oxford textbook of suicidology and suicide prevention: a global perspective*. Oxford: Oxford University Press (2009). p. 791–824.
18. David Klonsky E, Kotov R, Bakst S, Rabinowitz J, Bromet EJ. Hopelessness as a predictor of attempted suicide among first admission patients with psychosis: a 10-year cohort study. *Suicide Life-Threat Behav* (2012) 42(1):1–10. doi: 10.1111/j.1943-278X.2011.00066.x
19. McMillan D, Gilbody S, Beresford E, Neilly L. Can we predict suicide and non-fatal self-harm with the Beck Hopelessness Scale? A meta-analysis. *psychol Med* (2007) 37(6):769–78. doi: 10.1017/S0033291706009664
20. Kovacs M, Garrison B. Hopelessness and eventual suicide: a 10-year prospective study of patients hospitalized with suicidal ideation. *Am J Psychiatry* (1985) 142(5):559–63. doi: 10.1176/ajp.142.5.559
21. Kohout FJ, Berkman LF, Evans DA, Cornoni-Huntley J. Two shorter forms of the CES-D depression symptoms index. *J Aging Health* (1993) 5(2):179–93. doi: 10.1177/089826439300500202
22. Huth-Bocks AC, Kerr DC, Ivey AZ, Kramer AC, King CA. Assessment of psychiatrically hospitalized suicidal adolescents: self-report instruments as predictors of suicidal thoughts and behavior. *J Am Acad Adolesc Psychiatry* (2007) 46(3):387–95. doi: 10.1097/chi.0b013e31802b9535
23. Korkmaz H, Korkmaz S, Çakar M. Suicide risk in chronic heart failure patients and its association with depression, hopelessness and self esteem. *J Clin Neurosci* (2019) 68:51–4. doi: 10.1016/j.jocn.2019.07.062
24. Pompili M, Iliceto P, Lester D, Innamorati M, Girardi P, Tatarelli R. *Beck Hopelessness Scale-manuale*. Firenze: Organizzazioni Speciali (2008).
25. Pompili M, Rinaldi G, Lester D, Girardi P, Ruberto A, Tatarelli R. Hopelessness and suicide risk emerge in psychiatric nurses suffering from burnout and using specific defense mechanisms. *Arch Psychiatr Nurs* (2006) 20(3):135–43. doi: 10.1016/j.apnu.2005.12.002
26. Pompili M, Innamorati M, Gonda X, Serafini G, Sarno S, Erbutto D, et al. Affective temperaments and hopelessness as predictors of health and social functioning in mood disorder patients: a prospective follow-up study. *J Affect Disord* (2013) 150(2):216–22. doi: 10.1016/j.jad.2013.03.026
27. Hill RD, Gallagher D, Thompson LW, Ishida T. Hopelessness as a measure of suicidal intent in the depressed elderly. *Psychol Aging* (1988) 3(3):230. doi: 10.1037/0882-7974.3.3.230
28. Bouvard M, Charles S, Guerin J, Aimard G, Cottiaux J. Study of Beck's hopelessness scale. Validation and factor analysis. *L'Encephale* (1992) 18(3):237–40.
29. Steer RA, Kumar G, Beck AT. Hopelessness in adolescent psychiatric inpatients. *psychol Rep* (1993) 72(2):559–64. doi: 10.2466/pr0.1993.72.2.559
30. Dyce JA. Factor structure of the beck hopelessness scale. *J Clin Psychol* (1996) 52(5):555–8. doi: 10.1002/(SICI)1097-4679(199609)52:5<555::AID-JCLP10>3.0.CO;2-D
31. Aish A-M, Wasserman D. Does Beck's Hopelessness Scale really measure several components? *psychol Med* (2001) 31(2):367–72. doi: 10.1017/S0033291701003300
32. Hanna D, White R, Lyons K, McParland MJ, Shannon C, Mulholland C. The structure of the Beck Hopelessness Scale: A confirmatory factor analysis in UK students. *Pers Individ Differ* (2011) 51(1):17–22. doi: 10.1016/j.paid.2011.03.001
33. Ayub N. Measuring hopelessness and life orientation in Pakistani adolescents. *Crisis* (2009) 30(3):153–60. doi: 10.1027/0227-5910.30.3.153
34. Flamenbaum R, Delisle M, Holden R eds. *Factor structure of the Beck Hopelessness Scale among suicide attempters*. San Francisco: 115th Annual Convention of the American Psychological Association (2007).
35. Kao Y-C, Liu Y-P, Lu C-W. Beck hopelessness scale: exploring its dimensionality in patients with schizophrenia. *Psychiatr Quart* (2012) 83(2):241–55. doi: 10.1007/s11126-011-9196-9
36. Pompili M, Tatarelli R, Rogers JR, Lester D. The hopelessness scale: a factor analysis. *psychol Rep* (2007) 100(2):375–8. doi: 10.2466/pr0.100.2.375-378
37. Tanaka E, Sakamoto S, Ono Y, Fujihara S, Kitamura T. Hopelessness in a community population: Factorial structure and psychosocial correlates. *J Soc Psychol* (1998) 138(5):581–90. doi: 10.1080/00224549809600413
38. Steed L. Further validity and reliability evidence for Beck Hopelessness Scale scores in a nonclinical sample. *Educ psychol Meas* (2001) 61(2):303–16. doi: 10.1177/00131640121971121
39. Kruey PM, Emons WH, Sijtsma K. On the shortcomings of shortened tests: A literature review. *Int J Test* (2013) 13(3):223–48. doi: 10.1080/15305058.2012.703734
40. Balsamo M, Cataldi F, Carlucci L, Fairfield B. Assessment of anxiety in older adults: a review of self-report measures. *Clin Interv Aging* (2018) 13:573. doi: 10.2147/CIA.S114100
41. Balsamo M, Cataldi F, Carlucci L, Padulo C, Fairfield B. Assessment of late-life depression via self-report measures: a review. *Clin Interv Aging* (2018) 13:2021. doi: 10.2147/CIA.S178943
42. Higginson I. Measuring quality of life: Using quality of life measures in the NCBI. *BMJ* (2001) 322(7297):1297–300. doi: 10.1136/bmj.322.7297.1297
43. Cheung YB, Liu KY, Yip PS. Performance of the CES-D and its short forms in screening suicidality and hopelessness in the community. *Suicide Life-Threat Behav* (2007) 37(1):79–88. doi: 10.1521/suli.2007.37.1.79
44. Balsamo M, Saggino A, Carlucci L. Tailored screening for late-life depression: A short version of the Teate Depression Inventory (TDI-E). *Front Psychol* (2019) 10:2693. doi: 10.3389/fpsyg.2019.02693
45. Moran LA, Guyatt GH, Norman GR. Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *J Clin Epidemiol* (2001) 54(6):571–9. doi: 10.1016/S0895-4356(00)00342-5
46. Shrout PE, Yager TJ. Reliability and validity of screening scales: Effect of reducing scale length. *J Clin Epidemiol* (1989) 42(1):69–78. doi: 10.1016/0895-4356(89)90027-9
47. Carlucci L, Tommasi M, Balsamo M, Furnham A, Saggino A. Theology. Religious fundamentalism and psychological well-being: An Italian study. *J Psychol Theol* (2015) 43(1):23–33. doi: 10.1177/009164711504300103
48. Aloba O, Akinsulore A, Mapayi B, Oloniyi I, Mosaku K, Alimi T, et al. The Yoruba version of the Beck Hopelessness Scale: psychometric characteristics and correlates of hopelessness in a sample of Nigerian psychiatric outpatients. *Compr Psychiatry* (2015) 56:258–71. doi: 10.1016/j.comppsy.2014.09.024
49. Fraser L, Burnell M, Salter LC, Fourkala E-O, Kalsi J, Ryan A, et al. Identifying hopelessness in population research: a validation study of two brief measures of hopelessness. *BMJ Open* (2014) 4(5):e005093. doi: 10.1136/bmjopen-2014-005093
50. Weijters B, Baumgartner H, Schillewaert N. Reversed item bias: An integrative model. *psychol Methods* (2013) 18(3):320–34. doi: 10.1037/a0032121
51. de Ayala RJ. An introduction to polytomous item response theory models. *Meas Eval Couns Dev* (1993) 25(4):172–89.
52. Hambleton RK, Jones RW. An NCME instructional module on: Comparison of classical test theory and item response theory and their applications to test development. *Educ Meas: Issues Pract* (1993) 12(3):38–47. doi: 10.1111/j.1745-3992.1993.tb00543.x
53. Welch C, Hoover H. Procedures for extending item bias detection techniques to polytomously scored items. *Appl Meas Educ* (1993) 6(1):1–19. doi: 10.1207/s15324818ame0601\_1
54. Reise SP, Waller NG. Item response theory and clinical measurement. *Annu Rev Clin Psychol* (2009) 5:27–48. doi: 10.1146/annurev.clinpsy.032408.153553
55. Jabrayilov R, Emons WH, Sijtsma K. Comparison of classical test theory and item response theory in individual change assessment. *Appl psychol Meas* (2016) 40(8):559–72. doi: 10.1177/0146621616664046
56. Adedoyin OO, Nenty H, Chilisa B. Investigating the invariance of item difficulty parameter estimates based on CTT and IRT. *Educ Res Rev* (2008) 3(3):083–93.
57. Aggen SH, Neale MC, Kendler KS. DSM criteria for major depression: evaluating symptom patterns using latent-trait item response models. *psychol Med* (2005) 35(4):475–87. doi: 10.1017/S0033291704003563
58. Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of item response theory*. Newbury Park, CA: Sage (1991).

59. Sheehan D, Lecrubier Y. *The mini international neuropsychiatric interview version 6.0 (MINI 6.0)*. Jacksonville, FL: Medical Outcomes System Inc. (2010).
60. Innamorati M, Lester D, Balsamo M, Erbuto D, Ricci F, Amore M, et al. Factor validity of the Beck Hopelessness Scale in Italian medical patients. *J Psychopathol Behav Assess* (2014) 36(2):300–7. doi: 10.1007/s10862-013-9380-3
61. Mokken RJ. *A theory and procedure of scale analysis: With applications in political research*. New York: Walter-Gruyter-Mouton (1971).
62. Emons WH, Sijtsma K, Meijer RR. On the consistency of individual classification using short scales. *psychol Methods* (2007) 12(1):105–20. doi: 10.1037/1082-989X.12.1.105
63. Wind A. Examining rating scales using Rasch and Mokken models for rater-mediated assessments. *J Appl Meas* (2014) 15(2):100–32.
64. Edelsbrunner PA, Dablander F. The psychometric modeling of scientific reasoning: a review and recommendations for future avenues. *Educ Psychol Rev* (2019) 31(1):1–34. doi: 10.1007/s10648-018-9455-5
65. Sijtsma K, Molenaar IW. Mokken models. In W. J. van der Linden (Ed.), *Handbook of item response theory*. Boca Raton, FL: Chapman and Hall/CRC (2016) Volume One: Models p. 303–321.
66. Molenaar IW. *Nonparametric models for polytomous responses. Handbook of modern item response theory*. New York, NY: Springer (1997) p. 369–80.
67. Molenaar IW. Mokken scaling revisited. *Kwantitatieve Methoden* (1982) 3(8):145–64.
68. Wind S. Practice. An instructional module on Mokken scale analysis. *Educational Measurement: Issues and Practice* (2017) 36(2):50–66. doi: 10.1111/emip.12153
69. Engelhard G Jr. Historical perspectives on invariant measurement: Guttman, Rasch, and Mokken. *Measurement* (2008) 6(3):155–89. doi: 10.1080/15366360802197792
70. Van Schuur WH. Mokken scale analysis: Between the Guttman scale and parametric item response theory. *Political Analysis* (2003) 11(2):139–63. doi: 10.1093/pan/mpg002
71. Sijtsma K, Meijer RR, van der Ark LA. Mokken scale analysis as time goes by: An update for scaling practitioners. *Pers Individ Differ* (2011) 50(1):31–7. doi: 10.1016/j.paid.2010.08.016
72. van der Ark LA. New developments in Mokken scale analysis in R. *J Stat Software* (2012) 48(5):1–27. doi: 10.18637/jss.v048.i05
73. van der Ark LA. Mokken scale analysis in R. *J Stat Software* (2007) 20(11):1–19. doi: 10.18637/jss.v020.i11
74. Hemker BT, Sijtsma K, Molenaar IW. Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken I RT model. *Appl psychol Meas* (1995) 19(4):337–52. doi: 10.1177/014662169501900404
75. Straat JH, van der Ark LA, Sijtsma K. Using conditional association to identify locally independent item sets. *Methodology* (2016) 12:117–23. doi: 10.1027/1614-2241/a000115
76. Junker BW, Sijtsma K. Latent and manifest monotonicity in item response models. *Appl psychol Meas* (2000) 24(1):65–81. doi: 10.1177/01466216000241004
77. Ligtoet R, Van der Ark LA, te Marvelde JM, Sijtsma K. Investigating an invariant item ordering for polytomously scored items. *Educ psychol Meas* (2010) 70(4):578–95. doi: 10.1177/0013164409355697
78. Sijtsma K, Molenaar IW. Reliability of test scores in nonparametric item response theory. *Psychometrika* (1987) 52(1):79–97. doi: 10.1007/BF02293957
79. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* (1951) 16(3):297–334. doi: 10.1007/BF02310555
80. van der Ark LA, van der Palm DW, Sijtsma K. A latent class approach to estimating test-score reliability. *Appl psychol Meas* (2011) 35(5):380–92. doi: 10.1177/0146621610392911
81. Schoonjans F, Zalata A, Depuydt C, Comhaire F. MedCalc: a new computer program for medical statistics. *Comput Methods Progr Biomed* (1995) 48(3):257–62. doi: 10.1016/0169-2607(95)01703-8
82. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* (1982) 143(1):29–36. doi: 10.1148/radiology.143.1.7063747
83. Swets JA, Dawes RM, Monahan J. Better decisions through science. *Sci Am* (2000) 283(4):82–7. doi: 10.1038/scientificamerican1000-82
84. Sijtsma K, Molenaar IW. *Introduction to nonparametric item response theory*. Thousand Oaks: Sage (2002).
85. De Berardis D, Fornaro M, Orsolini L, Valchera A, Carano A, Vellante F, et al. Alexithymia and suicide risk in psychiatric disorders: a mini-review. *Front Psychiatry* (2017) 8:148. doi: 10.3389/fpsy.2017.00148
86. Serafini G, Lamis DA, Aguglia A, Amerio A, Nebbia J, Geoffroy PA, et al. Hopelessness and its correlates with clinical outcomes in outpatient setting. *J Affect Disord* (2019) 263:472–9. doi: 10.1016/j.jad.2019.11.144
87. Zhang J, Li Z. The association between depression and suicide when hopelessness is controlled for. *Compr Psychiatry* (2013) 54(7):790–6. doi: 10.1016/j.comppsy.2013.03.004
88. Nunnally J. *Psychometric theory*. (2nd ed.). New York: McGraw-Hill (1978).
89. Meijer RR, Egberink IJ. Investigating invariant item ordering in personality and clinical scales: Some empirical findings and a discussion. *Educ psychol Meas* (2012) 72(4):589–607. doi: 10.1177/0013164411429344
90. Smith GT, McCarthy DM, Anderson KG. On the sins of short-form development. *Psychol Assess* (2000) 12(1):102. doi: 10.1037/1040-3590.12.1.102
91. Credé M, Harms P, Niehorster S, Gaye-Valentine AJ. An evaluation of the consequences of using short measures of the Big Five personality traits. *J Pers Soc Psychol* (2012) 102(4):874. doi: 10.1037/a0027403
92. Smith ST, Edens JF, Vaughn MG. Assessing the external correlates of alternative factor models of the Psychopathic Personality Inventory–Short Form across three samples. *J Pers Assess* (2011) 93(3):244–56. doi: 10.1080/00223891.2011.558876
93. Niméus A, Träskman-Bendz L, Alsén MJ. Hopelessness and suicidal behavior. *J Affect Disord* (1997) 42(2-3):137–44. doi: 10.1016/S0165-0327(96)01404-8
94. Reise SP, Waller NG. How many IRT parameters does it take to model psychopathology items? *psychol Methods* (2003) 8(2):164–84. doi: 10.1037/1082-989X.8.2.164
95. Chen Y, Watson R, Hilton A. An exploration of the structure of mentors' behavior in nursing education using exploratory factor analysis and Mokken scale analysis. *Nurse Educ Today* (2016) 40:161–7. doi: 10.1016/j.nedt.2016.03.001
96. Smits IA, Timmerman ME, Meijer RR. Exploratory Mokken Scale Analysis as a dimensionality assessment tool: Why scalability does not imply unidimensionality. *Appl psychol Meas* (2012) 36(6):516–39. doi: 10.1177/0146621612451050
97. Meijer RR, Baneke JJ. Analyzing psychopathology items: a case for nonparametric item response theory modeling. *psychol Methods* (2004) 9(3):354. doi: 10.1037/1082-989X.9.3.354
98. Classen S, Wang Y, Winter SM, Velozo CA, Lanford DN, Bédard M. Concurrent criterion validity of the safe driving behavior measure: A predictor of on-road driving outcomes. *Am J Occup Ther* (2013) 67(1):108–16. doi: 10.5014/ajot.2013.005116

**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.

The handling editor is currently co-organizing a Research Topic with one of the authors, MP, and confirms the absence of any other collaboration.

Copyright © 2020 Balsamo, Carlucci, Innamorati, Lester and Pompili. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.