



Psychometric Study Using Item Response Theory of an Instrument Developed for Assessment of Iranian Mental Health Problems

Masoumeh Dejman ¹, Monir Baradaran Eftekhari ², Katayoun Falahat ², Zohreh Mahmoodi ³, Mojgan Padyab ⁴ and Ameneh Setareh Forouzan ^{5,*}

¹Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

²Deputy for Research and Technology, Ministry of Health and Medical Education, Tehran, Iran

³Social Determinants of Health Research Center, Alborz University of Medical Sciences, Karaj, Iran

⁴Department of Social Work, Umeå University, Umeå, Sweden

⁵Social Welfare Management Research Center, University of welfare and Rehabilitation, Tehran, Iran

*Corresponding author: Social Welfare Management Research Center, University of welfare and Rehabilitation, Tehran, Iran. Email: asforouzan@gmail.com

Received 2021 July 11; Revised 2022 May 07; Accepted 2022 July 28.

Abstract

Background: Currently, in addition to the undeniable impact of cultural factors on mental health problems' diagnosis and treatment methods, the use of rapid, short, and intervention-based instruments can be effective in the accurate diagnosis of mental health problems, especially in the health system of developing countries.

Objectives: This study aimed to validate an instrument developed for screening patients with common mental health problems using item response theory (IRT).

Methods: The study was conducted in Semnan province (with Persian ethnicity), Iran, from August 2017 to February 2018. A 101-item tool consisted of district common mental health problems (i.e., depression, anxiety, and obsession), along with a functional checklist. The development of the instrument involved a pilot study and psychometric testing. The IRT-based analysis was used as the item-reduction method to evaluate the shortened tool as an appropriate screening tool. The participants were healthy individuals and patients with depression, anxiety, and obsessive-compulsive disorder (OCD). The data were analyzed using Stata software (version 15.1).

Results: The study participants were 160 individuals (58.2% male) with a mean age of 36.3 ± 11.2 years. All item impact factors were within the range of 1.8 - 5. The mean values of clarity, simplicity, relevance, and scale-level content validity index/averaging calculation method of the instrument were 96.73 ± 0.70 , 97.64 ± 0.61 , 98.2 ± 1.9 , and 97.09 ± 0.63 , respectively. Cronbach's alpha and internal consistency coefficient were 0.88 and 0.7. Moreover, 13, 5, and 12 items were excluded using IRT from depression, anxiety, and OCD dimensions based on the threshold criteria, respectively.

Conclusions: Iranian screening tools for mental health problems can provide qualified information with the least error and the most precision in appropriate early diagnosis and decrease the burden of mental health problems in the national healthcare system.

Keywords: Anxiety Disorder, Depression, Item Response Theory, Obsessive-Compulsive Disorder, Validity, Reliability

1. Background

In recent years, the burden of mental health problems, with their significant effects on health, socioeconomic consequences, and human rights in all countries of the world, has been increasing. According to studies, 1 in every 5 individuals suffers from some kind of mental health problem (1). However, lack of appreciation of the need for treatment and financial and other barriers hindering access to it, stigmas associated with mental health problems, disadvantages experienced by minorities, and cultural differences impede early diagnosis and treatment (2), which can lead to increased mental health treatment

gap due to limitations in evidence-based mental health intervention resources (3).

Cultural difference is one of the diagnostic and therapeutic barriers to mental health problems (4, 5). This factor has effects on an individual's understanding of the disease, the need for treatment, and health-seeking behaviors, such as attitudes toward therapeutic and preventive care (3). Cultural and environmental differences cause individuals to describe and prioritize their psychosocial problems in different ways, which can affect receiving therapeutic interventions (6). Therefore, diagnostic-therapeutic interventions in every country should be tailored to the cultural specifics of that particular country (7). Diagnostic

and measurement errors are other factors that can arise and cause discrepancies in this regard (8).

A variety of diagnostic and screening instruments are available for the assessment of mental health problems and their interventions (9). However, few of these instruments are specifically designed to evaluate interventions in developing countries, which is one of the concerns and causes of patient neglect in these countries (6). Instruments used for disease screening and diagnosis in Western societies might not be able to detect the symptoms of mental health problems in other cultures, such as the Iranian society, because individuals in these cultures might consider some symptoms as usual and not as indicators of disease (1). Another factor determining the appropriateness of an instrument is the quality of being patient-centered to not impose a burden on the patient in terms of the number of items or complexity of the questionnaire (10). Short questionnaires are more appealing because they are easy to use and time-saving and decrease the burden of responsibility, thereby minimizing information loss (11). Designing shorter instruments is relatively common in all areas of psychology and psychiatry, and it is also desirable to reduce the number of items in the existing questionnaires (12). Several approaches are used in reducing the number of items, including the approaches of concept-retention, the equi-discriminative item-total correlation, and the Rasch model (4).

As part of a larger study on the assessment of culturally appropriate interventions, this paper describes the development of an instrument, based on item response theory (IRT), for the proper assessment of interventions. The IRT method is used to assess the cross-population utility of instruments that models the probability of a latent variable based on a given response as a function of the respondents. The IRT methodology allows how a tool performs in various populations by presenting information on identifying potential item-level bias and item characteristics (3).

2. Objectives

This study was part of the main intervention program on common mental problems. The present study aimed to validate the instrument developed for screening patients with common mental health problems using IRT. The current study was conducted on the adult population in two cities in Semnan province, east of Tehran (the capital), in northern-central Iran. The location was selected due to its proximity to the capital of the country and its low rate of immigration. According to the census of 2016, the population of the province was 352,285 individuals, comprising 36,298 households. In a national survey in 2017, the prevalence of suspected cases of mental health problems

was 14.5% (13.1% of male and 15.8% of female individuals) (6). About 20%, 23.8%, and 7.2% of the sample had somatization (13.5% of male and 21.4% of female individuals), anxiety (17.7% of male and 26.8% of female individuals), and depression, respectively (6).

3. Methods

3.1. Preceding Qualitative and Instrument Validation Studies

A qualitative study was conducted to identify local mental health problems through developing, adapting, and validating a native instrument. The qualitative study showed a wide range of internalizing and externalizing symptoms (13). It was shown that the community was mostly concerned about symptoms related to common mental problems of depression, anxiety, and obsessive-compulsive disorder (OCD). All the existing validated measurement instruments for these three common mental disorders were listed and reviewed by the authors, and three instruments that were more symptom-based were selected. Prioritized symptoms/tasks or words/idioms selected in the qualitative study not present in these standard instruments were added with minor adjustments. For function assessment, a checklist, based on the qualitative findings of the study, was developed and reviewed for face validity with a group of the target population. Another review was performed with the interviewers as part of their training courses before the instrument test.

3.2. Measures

The following instruments were selected and modified:

3.2.1. Yale-Brown Obsessive-Compulsive Scale

Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) assesses the symptoms of OCD in terms of severity. Wayne K. Goodman et al. created this scale, which is widely used in research and treatment of patients to assess the severity of OCD symptoms and the response to treatment. This scale has 90 items and rates obsessions and compulsions separately (14). The mean reliability values of the total scale are 0.866, 0.922, and 0.848 for the alpha coefficient, inter-class correlations, and test-retest correlations, respectively (9). In the current study, some questions were modified or added based on the preceding qualitative findings. The modified instrument contained 21 questions.

3.2.2. Beck Depression Inventory Scale

The Beck Depression Inventory (BDI) scale was designed by Aaron T. Beck with 21 questions and is self-scored. This psychometric test is extensively used to assess the

severity of depression (15). The modified instrument included 28 questions (1). The questions regarding religious beliefs, social networks, and digestive problems were added to the original questionnaire. In addition, the wording of some questions was updated according to the results of the qualitative research, including adding the sentence “I am locked-in due to too much thought” to Q10 and adding the word “penalty” to Q14 (1).

3.2.3. Beck Anxiety Inventory Scale

The Beck Anxiety Inventory (BAI) scale was designed by Aaron T. Beck et al. It is an inventory to determine the degree of anxiety in adults and children and is self-reported (16, 17). The modified instrument had 27 questions. Seven questions concerning aggression, sleep disorders, impatience/haste, disturbing thoughts, lack of focus, and worry were added to the BAI.

3.2.4. Function Checklist

This instrument was developed based on the results of the qualitative research with 25 questions. This checklist assesses different aspects of the individual (e.g., exercise, study, and recreation), familial (e.g., taking care of spouse/children and doing housework), and social (e.g., voluntary activities or helping others) activities. The numbers of questions related to individual, familial, and social aspects were 15, 6, and 4, respectively.

3.2.5. Final Package

The final version of the instrument consisted of 101 questions (depression: 1 - 28, anxiety: 29 -55, obsession: 56 - 76, and function: 77 - 101). The psychometrics of the final instrument was evaluated using face validity, content validity, criterion validity, and reliability as follows:

3.3. Psychometric Characteristics

3.3.1. Face Validity

Qualitative and quantitative methods were used to determine the face validity of the questionnaire. Face-to-face interviews were held with 10 members of the general public to determine the qualitative face validity. Items' relevance and relationships and ambiguities, and difficulties in understanding the concepts and words were assessed. Modifications were then made according to participants' views. The item impact score method was used to decrease the number of items, omit the inappropriate ones, and determine their significance. The items were rated using a 5-point Likert scale (from “not important at all” to “very important”) by the respondents in the qualitative section, and the item impact score was separately measured for each item using the following equation:

$$\text{Item impact} = \text{frequency (\%)} \times \text{importance}$$

The items with an impact score ≥ 1.5 were deemed appropriate and were kept for the next stage of the analysis.

3.3.2. Content Validity

The content validity of the designed questionnaire was assessed using both qualitative and quantitative methods. The qualitative content validity assessment was performed based on 10 experts' feedback on the importance, necessity, item allocation, grammar, wording, and proper scoring of the scale. The content validity index (CVI) and content validity ratio (CVR) were determined subsequently. The copies of the designed questionnaire were distributed among 10 psychiatrists to assess each item considering a 3-point scale (“necessary”, “useful but not necessary”, and “unnecessary”). Based on scores obtained from Lawshe's table, based on the 10 experts' assessment, the item was deemed necessary and essential at the statistically significant level of $\alpha < 0.05$ (106) whether the minimum value of the index was greater than 0.62. An equation was used to find the CVR. Accordingly, the researcher distributed copies of the questionnaire among the 10 experts with the request to determine the simplicity, relevance, and clarity of the items based on Waltz and Bausell's validity index. The experts assessed all the items using a 4-point Likert scale (e.g., relevancy, from 1 = irrelevant to 4 = totally relevant).

A CVI greater than 0.79 indicated an appropriate item. A CVI within 0.7 - 0.79 indicated a debatable item requiring modification, and a CVI less than 0.7 indicated an unacceptable item that had to be eliminated (15). Then, the mean scale-level content validity index/averaging calculation method (S-CVI/Ave) was utilized based on the average CVI of all items. Polit and Beck have suggested an acceptability score of 90% and higher for the S-CVI/Ave (1).

3.3.3. Criterion Validity and Item Response Theory

This study was performed within August 2017 to February 2018. Two methods were used to recruit the participants. The subjects were selected from the urban health centers and referrals to the psychiatric center in the main hospital in Semnan. The inclusion criteria were age over 18 years, residence in Semnan province for at least 5 years, registration in the Integrated Health System (SIB), and informed consent to participate in the study. The exclusion criteria were inability to communicate and acute psychiatric symptoms.

For comparing the mean scores for each problem of those who had the problem to those who did not (arbitrarily use of 20% difference as the cut-off), 40 subjects were selected for each problem (a total of 120 participants) and 40

subjects without any problem or healthy. The project staff used the community health records repository (SIB online platform) to assemble the list of participants. In the health centers or psychiatric centers, the study interviewers explained the project to the participants, completed the informed consent process, and then administered the study instrument. After completing the instrument, the participants were referred to a trained psychiatrist to be assessed according to the Structured Clinical Interview for DSM Disorders (SCID) measurement.

3.4. Background Variables

Background variables were included in the sociodemographic section of the survey instrument. Age was measured in years. The educational level had three categories, namely primary (less than 5 years of education), intermediate (5 - 12 years of education), and high (more than 12 years of education).

3.5. Statistical Analysis

For the achievement of the overall aim of the study, we developed a short, reliable, and valid measure of three main mental health problems, namely anxiety, depression, and OCD, by shortening a questionnaire, including 28, 21, and 27 items for depression, OCD, and anxiety, respectively. The latent traits (i.e., depression, OCD, and anxiety) were measured using an instrument, which is a collection of items. Each item has a difficulty parameter and a discrimination parameter. The correct answer is given by the gold standard criterion using the SCID-based clinical assessment for each mental disorder. This study investigated the probability of a positive response to each item separately.

The IRT for binary outcome was used as the item-reduction method on cross-sectional field-tested data derived from 160 participants from healthcare centers in Semnan. A series of binary response items for each dimension was created in which the response option "Not at all" was regarded as 0, and response options 1-3 were regarded as 1. This study reported the discrimination parameter (a) and difficulty parameter (b) using the IRT parameterization. The items with high discrimination, based on thresholds reported in the literature, were retained in the final version of the questionnaire. Generally, item discrimination was regarded as very low to very high (Baker, 2001).

Reliability analysis was performed on the remaining items for each dimension through Cronbach's alpha and item-rest correlations. Sensitivity and specificity analyses were performed using the receiver operating characteristic curve to assign a cut-off point. Statistical analyses were carried out using Stata software (version 15.1; StataCorp, College Station, TX, USA).

3.6. Ethical Considerations

The study protocol was approved by the Research Ethics Committee of the National Institute for Medical Research Development (NIMAD) under the code IR.NIMAD.REC.1395.047. All the participants signed the consent form and verbally consented to participate in the study.

4. Results

There were 160 participants in the present study, 58.2% of whom were male. The mean age of the participants was 36.3 ± 11.2 years (range: 17 - 75). More than 50% of the participants had an intermediate education (Table 1). The results are presented in the two following parts:

Table 1. Descriptive Statistics of the Study Participants

Characteristics	Total (N = 160), No. (%)
Gender	
Male	93 (58.2)
Female	67 (36.4)
Marital status	
Married	122 (76.2)
Single	29 (18.2)
Widowed/Divorced	9 (5.6)
Educational level	
Primary	30 (18.8)
Intermediate	89 (55.6)
High	41 (25)

4.1. Validity of the Common Mental Health Screening Instrument

The initial common mental health scale package consisted of 101 questions (depression: 1 - 28; anxiety: 29 - 55; obsession: 56 - 76; function: 77 - 101). Face validity, content validity, and criterion validity were used to validate the instrument. In the face validity stage, the necessary changes were made after receiving the opinions of the members of the public. The impact factors of the items were within the range of 1.8 - 5, and none of the items had an impact factor of less than 1.5. Therefore, all the items were transferred to the content validity stage.

In the qualitative content validity stage, the necessary changes were made after receiving the opinions of experts in relevant fields. The CVR and CVI results were obtained. For all of the items, CVR and S-CVI/Ave were within the range of 0.7-1 and 97.09 ± 0.63 , respectively. The mean values of clarity, relevance, S-CVI/Ave, and simplicity were

96.73 ± 0.70, 98.2 ± 1.9, 97.09 ± 0.63, and 97.64 ± 0.61, respectively. In the criterion validity stage, the final package was examined with the SCID (Table 2).

Table 2. Comparison of Diagnosis of Structured Clinical Interview for DSM Disorders with Mean Scores of Subscales of Iranian Mental Health Problems

SCID	Depression Score	Anxiety Score	OCD Score
OCD (n = 40)	27.6 ± 13.65	28.89 ± 14.93	24.40 ± 13.47
Anxiety (n = 40)	22.24 ± 11.17	28.32 ± 11.32	15.76 ± 10.62
Depression (n = 40)	32.23 ± 10.78	32.8 ± 12.53	23 ± 11.84
Healthy (n = 40)	10.96 ± 6.48	7.37 ± 7.78	7.16 ± 5.17
Total (n = 160)	23.25 ± 14.03	24.35 ± 16.09	18.34 ± 13.47

Abbreviations: SCID, Structured Clinical Interview for DSM Disorders; SD, standard deviation; OCD, obsessive-compulsive disorder.

^a Values are expressed as mean ± SD.

4.2. Item Reduction

Table 3 shows the results for IRT, including discrimination and difficulty parameters. Discrimination parameters for the depression, OCD, and anxiety dimensions were within the ranges of $\alpha = 0.72$ (depression 27) to $\alpha = 2.59$ (depression 9), $\alpha = 0.95$ (OCD74) to $\alpha = 2.48$ (OCD70), and $\alpha = 1.11$ (anxiety 44) to $\alpha = 4.55$ (anxiety 54), respectively. Based on the threshold criteria, this study excluded items 3, 8, 10, 14, 15, 18, 19, 20, 21, 22, 25, 27, and 28 from the depression dimension, items 29, 30, 44, 46, and 51 from the anxiety dimension, and items 56, 57, 60, 61, 63, 64, 66, 68, 72, 73, 74, and 75 from the OCD dimension. Cronbach's alpha coefficients were calculated at 0.92, 0.95, and 0.88 for depression, anxiety, and OCD, respectively, suggesting high reliability for all three dimensions.

Sensitivity and specificity analyses were performed to assign a cut-off point. The choice for depression was based on 89% sensitivity and 59% specificity, determining a cut-off point of 14. This finding means that if the sum of the scores on items for the depression dimension exceeds 14, the individual is considered positive in the screening test. Similarly, considering 80% sensitivity and 44% specificity, a cut-off point of 17 was chosen for anxiety. Given 80% sensitivity and 49% specificity, a cut-point of 6 was considered for OCD (Table 4). In this study, Cronbach's alpha and internal consistency coefficients were 0.88 and 0.7, respectively.

5. Discussion

This study to develop and test an instrument is the first in a series of investigations to assess the performance of screening instruments. This series of studies includes vast

epidemiological studies and clinical validations to determine shared constructs between instruments using factor analysis to enhance the assessment of the screening and treatment of mental health problems in Semnan province. This study performed a transcultural translation of the BAI, BDI, and Y-BOCS. This study developed a function assessment instrument locally and learned through the process of qualitative transcultural translation that some items of the BDI and BAI were not applicable in the Persian version of this questionnaire due to the lack of interpersonal interpretation, specificity, or conceptual nonequivalence. In addition, it was understood that various interpretations of somatic idioms affected the proper recognition of the physical problems in mental health screening. Several crucial lessons can be learned from the adaptation and development of screening instruments during the process of expanding healthcare services to include mental health. It has been proposed that to correctly identify those individuals who require mental healthcare, it is essential to recognize practices and beliefs that are considered normal in a particular culture, such as communicating with deceased ancestors, and differentiate them from probable mental illness (12, 18-20).

In the present study, the final mental health scale package was developed as a short, reliable, and valid measure of three main mental health problems consisting of anxiety, depression, and OCD by shortening a questionnaire containing 28, 21, and 27 items for depression, OCD, and anxiety, respectively. Content and face validity studies confirmed the clarity and simplicity of the items. The relevance of the items in the CVI showed a significant degree of agreement among the experts.

The correlation coefficients of the values obtained from this instrument and the SCID were used for the criterion validation. The results presented a correlation between these instruments that measured the same issue.

Exploratory factor analysis was used to reduce the number of items in several studies (21). Specifically, this approach has been utilized to identify items with loadings below 0.4 on any conceived factor. This identification results in the elimination of the pertinent items from the model. The use of this method is limited by the fact that it accounts for neither the structure of the original factors nor the model's structure. Moreover, exploratory factor analysis is intended to be used to explore novel constructs and not existing scales (22). The IRT method is a kind of latent variable analysis utilized for a better understanding of how an instrument performs in various populations by identifying potential item-level bias and providing information on item characteristics across populations. The IRT analysis includes discrimination and difficulty parameters (3).

Table 3. Item Discrimination Parameter (A), Difficulty Parameter (B), and Standard Errors of Subscales of Iranian Mental Health Problems^a

	Depression			Anxiety			OCD	
	A (SE)	B (SE)		A (SE)	B (SE)		A (SE)	B (SE)
D1 Sadness	1.91 (0.37)	-0.69 (0.14)	A29 Numbness	1.13 (0.23)	-0.54 (0.18)	OCD56 Harming	1.47 (0.29)	0.60 (0.16)
D2 Crying	2.53 (0.46)	-0.02 (0.11)	A30 Feeling hot	1.65 (0.31)	-0.48 (0.14)	OCD57 Aggressive behaviors	1.68 (0.31)	-0.26 (0.13)
D3 Discourage	1.50 (0.30)	-0.91 (0.18)	A31 Wobbliness in leg	1.86 (0.34)	0.34 (0.13)	OCD58 Najes (Unclean)	1.63 (0.30)	0.18 (0.13)
D4 Pessemism	1.97 (0.36)	-0.39 (0.13)	A32 Unable to relax	3.56 (0.72)	-0.83 (0.12)	OCD59 Getting ill	1.93 (0.35)	0.09 (0.12)
D5 Self dislike	1.71 (0.37)	-1.4 (0.23)	A33 Worst happening	2.43 (0.46)	-0.80 (0.14)	OCD60 Cleaniness	1.58 (0.32)	-0.96 (0.18)
D6 Self-blame	2.32 (0.44)	-0.49 (0.12)	A34 Feeling dizzy	2.54 (0.46)	-0.22 (0.11)	OCD61 Forbidden sexual thoughts	1.19 (0.26)	0.77 (0.20)
D7 Worthless	2.00 (0.41)	-1.09 (0.18)	A35 Heart pounding	2.12 (0.38)	-0.36 (0.12)	OCD62 Store things	2.03 (0.37)	0.23 (0.12)
D8 Irritability	1.65 (0.33)	-0.85 (0.17)	A36 Unsteady mode	2.84 (0.54)	-0.70 (0.12)	OCD63 Blasphemous	1.19 (0.25)	0.70 (0.19)
D9 Boring	2.59 (0.56)	-1.13 (0.16)	A37 Terrified	2.31 (0.41)	-0.01 (0.11)	OCD64 Ethical issues	1.62 (0.31)	-0.60 (0.15)
D10 Busy minded	1.56 (0.35)	-1.51 (0.26)	A38 Nervous	3.29 (0.68)	-1.01 (0.14)	OCD65 Recalling things	2.17 (0.42)	-0.62 (0.13)
D11 Indecisiveness	2.38 (0.47)	-0.71 (0.13)	A39 Chocking	2.08 (0.38)	0.32 (0.12)	OCD66 Losing belongings	1.70 (0.32)	-0.38 (0.14)
D12 Concentrate	2.13 (0.42)	-0.80 (0.14)	A40 Hand trembling	1.84 (0.33)	0.21 (0.12)	OCD67 Inherent compulsion	2.23 (0.42)	-0.44 (0.12)
D13 Guilty	2.58 (0.47)	-0.05 (0.11)	A41 Body trembling	1.88 (0.35)	0.51 (0.13)	OCD68 Excessive care	1.59 (0.31)	-0.79 (0.17)
D14 Punishing	1.39 (0.27)	-0.26 (0.15)	A42 Losing control	2.79 (0.51)	-0.23 (0.11)	OCD69 Repeating daily activities	1.80 (0.33)	-0.19 (0.13)
D15 Feeling in trouble	1.07 (0.25)	-1.26 (0.29)	A43 Difficulty in breathing	2.36 (0.42)	0.14 (0.11)	OCD70 Unpleasant thoughts	2.49 (0.46)	-0.70 (0.13)
D16 Suicide	1.90 (0.37)	0.79 (0.15)	A44 Fear of dying	1.11 (0.23)	0.25 (0.17)	OCD71 On time	1.92 (0.36)	-0.40 (0.13)
D17 Death thoughts	1.52 (0.30)	-0.52 (0.15)	A45 Scared	2.07 (0.37)	-0.20 (0.12)	OCD72 Superstitious	1.38 (0.27)	0.59 (0.17)
D18 Appearance	1.23 (0.25)	-0.46 (0.17)	A46 Indigestion	1.07 (0.22)	-0.21 (0.17)	OCD73 Personal hygiene	1.46 (0.28)	0.23 (0.14)
D19 No interest	1.51 (0.28)	0.21 (0.14)	A47 Faint	1.42 (0.38)	2.04 (0.40)	OCD74 Bathing time	0.95 (0.21)	0.26 (0.19)
D20 Weight	0.86 (0.21)	-1.24 (0.33)	A48 Face flushed	1.77 (0.32)	0.09 (0.12)	OCD75 Several recheck	1.51 (0.28)	-0.09 (0.14)
D21 Less sex	0.85 (0.21)	-0.65 (0.24)	A49 Sweat	1.76 (0.32)	-0.05 (0.12)	OCD76 Repeating words	2.32 (0.43)	-0.31 (0.12)
D22 Appetite	1.40 (0.28)	-0.67 (0.17)	A50 Aggression	2.23 (0.41)	-0.79 (0.14)			
D23 Sleep	2.12 (0.41)	-0.62 (0.13)	A51 Sleep problem	1.62 (0.31)	-0.86 (0.17)			
D24 Effort	2.30 (0.44)	-0.57 (0.13)	A52 Impatience	2.11 (0.39)	-0.67 (0.14)			
D25 Indigestion	1.07 (0.23)	-0.30 (0.17)	A53 Worrying thoughts	3.21 (0.62)	-0.53 (0.11)			
D26 Relationship	2.55 (0.47)	-0.21 (0.11)	A54 Worried	4.55 (1.05)	-0.74 (0.11)			
D27 Social network	0.72 (0.20)	0.49 (0.26)	A55 Lack of concentration	2.99 (0.57)	-1.68 (0.12)			
D28 Beleif	0.82 (0.20)	0.39 (0.22)						

Abbreviations: OCD, obsessive-compulsive disorder; SE, standard errors.

^a A, item discrimination parameter; B, difficulty parameter.

Table 4. Item Correlations and Alpha Coefficients of Subscales of Iranian Mental Health Problems

Depression					Anxiety					OCD				
Item	Item-Test Correlation	Item-Rest Correlation	Interitem Covariance	Alpha	Item	Item-Test Correlation	Item-Rest Correlation	Interitem Covariance	Alpha	Item	Item-Test Correlation	Item-Rest Correlation	Interitem Covariance	Alpha
Dep1	0.699	0.638	0.418	0.913	Anx31	0.641	0.603	0.447	0.95	OCD58	0.711	0.621	0.489	0.875
Dep2	0.704	0.646	0.42	0.913	Anx32	0.775	0.745	0.433	0.95	OCD59	0.675	0.576	0.497	0.876
Dep4	0.642	0.581	0.43	0.915	Anx33	0.753	0.72	0.434	0.95	OCD62	0.688	0.602	0.501	0.876
Dep5	0.709	0.664	0.43	0.912	Anx34	0.695	0.69	0.442	0.95	OCD65	0.751	0.671	0.48	0.87
Dep6	0.716	0.657	0.416	0.912	Anx35	0.66	0.618	0.441	0.95	OCD67	0.783	0.712	0.471	0.867
Dep7	0.647	0.598	0.43	0.914	Anx36	0.733	0.699	0.437	0.95	OCD69	0.665	0.564	0.499	0.879
Dep9	0.727	0.677	0.421	0.912	Anx37	0.704	0.666	0.438	0.95	OCD70	0.751	0.675	0.483	0.87
Dep11	0.705	0.647	0.418	0.913	Anx38	0.76	0.728	0.433	0.95	OCD71	0.715	0.622	0.484	0.875
Dep12	0.639	0.577	0.431	0.915	Anx39	0.635	0.596	0.448	0.95	OCD76	0.77	0.687	0.466	0.869
Dep13	0.734	0.683	0.419	0.911	Anx40	0.652	0.616	0.448	0.95	Test Scale			0.485	0.886
Dep16	0.596	0.536	0.44	0.916	Anx41	0.65	0.617	0.45	0.95					
Dep17	0.646	0.581	0.429	0.915	Anx42	0.745	0.714	0.438	0.95					
Dep23	0.668	0.603	0.422	0.914	Anx43	0.671	0.638	0.447	0.95					
Dep24	0.704	0.653	0.425	0.912	Anx45	0.673	0.633	0.441	0.95					
Dep26	0.724	0.673	0.421	0.912	Anx47	0.389	0.358	0.471	0.95					
Test Scale			0.425	0.919	Anx48	0.616	0.577	0.45	0.95					
					Anx49	0.621	0.578	0.445	0.95					
					Anx50	0.725	0.69	0.436	0.95					
					Anx52	0.728	0.688	0.431	0.95					
					Anx53	0.791	0.761	0.428	0.95					
					Anx54	0.822	0.796	0.426	0.94					
					Anx55	0.741	0.707	0.434	0.95					
					Test Scale			0.441	0.95					

Abbreviations: OCD, obsessive-compulsive disorder.

The study of any subject needs a suitable tool for data collection with the least number of errors and the most level of precision (23). An instrument designed in a country is influenced by the culture of that country, and its use elsewhere, even via precise translation, might lead to problems due to the content considered not being a good fit (24). In short, psychometrically sound measures introduce more efficient methods of quantifying patient outcomes to researchers and clinicians while retaining the validity and reliability of the longer versions. Such tools offer the advantage of providing the same quality of information with less burden for the patient and easier scoring for the researcher or clinician (4).

The comparison of the results of the present study to other similar studies indicates that this tool has a perfect sensitivity to identify mental health problems. Brief Jail Mental Health Screen generally had a sensitivity of approximately 65% (% 95 CI: 47-48%) (25). England Mental Health Screen has only 50% sensitivity in a small subsample of 18-21-year-old male subjects (26). In the current study, the choice for depression, anxiety, and OCD was 80-89% sensitivity and 44 - 59% specificity.

In this study, Cronbach alpha coefficients were calcu-

lated at 0.92, 0.95, and 0.88 for depression, anxiety, and OCD, respectively. Therefore, it presented a high internal consistency among the items confirming the reliability of the instrument. However, the reliability coefficient of the General Health Questionnaire based on Cronbach alpha was calculated at 90 (27).

A few challenges encountered in the studies warrant discussion. Firstly, local supervisors had difficulty finding and/or finishing a pilot case. Feedback received from all supervisors suggested that time was the primary barrier to not completing a case. Other barriers to the projects across both cities were mainly organizational and logistical, such as transport and personnel problems. Future studies are needed to examine organizational facilitators and barriers. With regard to the duration of follow-up, this study did not evaluate the longitudinal effects of the intervention. Although the follow-up was, on average, longer than one month after treatment, additional postintervention assessments of 6-12 months after treatment would be more informative.

5.1. Conclusions

Iranian mental health problems screening tools can provide qualified information with the least error and

the most precision in appropriate early diagnosis and decrease the burden of mental health problems in the national healthcare system. This questionnaire has 71 questions, including 15, 9, and 22 items for depression, OCD, and anxiety, respectively, and one checklist with 25 items for the function.

The role of the lack of mental health professionals (i.e., psychiatrists or psychologists) shifting from treating a few cases to supervising the treatment of many individuals through the community health workers is supported as community health workers were able to learn and provide both interventions with fidelity. This approach to task sharing is supported by other studies as a good option for providing sustainable, accessible, and effective services for multiple mental health problems at a scale where there are few professionals.

Footnotes

Authors' Contribution: MD supervised the study and data analysis and wrote the manuscript. MBE contributed to the study design and wrote the manuscript. KF contributed to the study design and collected the data. ZM contributed to the study design, analyzed the data, and drafted the manuscript. MP contributed to data analysis. ASF contributed to the study design, supervised the study, and wrote the manuscript.

Conflict of Interests: The authors declare no potential conflict of interest.

Data Reproducibility: The data presented in this study are openly available in one of the repositories or will be available on request from the corresponding author by this journal representative at any time during submission or after publication. Otherwise, all the consequences of possible withdrawal or future retraction will be with the corresponding author.

Ethical Approval: The study protocol was approved by the Research Ethics Committee of the NIMAD under the code IR.NIMAD.REC.1395.047.

Funding/Support: This study was financially supported partly by grant number 943114 from the NIMAD.

Informed Consent: All the participants signed the consent form and verbally consented to participate in the study.

References

- Baradaran Eftekhari M, Dejman M, Forouzan AS, Falahat K, Shati M, Mirabzadeh A, et al. Developing a Depression Inventory for Screening the Fars Ethnicity in Iran. *Iran J Psychiatry Behav Sci.* 2019;**13**(3). doi: 10.5812/ijpbs.82646.
- Sexton E, King-Kallimanis BL, Morgan K, McGee H. Development of the brief ageing perceptions questionnaire (B-APQ): a confirmatory factor analysis approach to item reduction. *BMC Geriatr.* 2014;**14**:44. doi: 10.1186/1471-2318-14-44. [PubMed: 24716631]. [PubMed Central: PMC4021231].
- Haroz EE, Bolton P, Gross A, Chan KS, Michalopoulos L, Bass J. Depression symptoms across cultures: an IRT analysis of standard depression symptoms using data from eight countries. *Soc Psychiatry Psychiatr Epidemiol.* 2016;**51**(7):981-91. doi: 10.1007/s00127-016-1218-3. [PubMed: 27083900]. [PubMed Central: PMC6022281].
- Beaton DE, Wright JG, Katz JN, Upper Extremity Collaborative G. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am.* 2005;**87**(5):1038-46. doi: 10.2106/JBJS.D.02060. [PubMed: 15866967].
- Nejatian M, Tehrani H, Momeniyan V, Jafari A. A modified version of the mental health literacy scale (MHLS) in Iranian people. *BMC Psychiatry.* 2021;**21**(1):53. doi: 10.1186/s12888-021-03050-3. [PubMed: 33485306]. [PubMed Central: PMC7824912].
- Noorbala AA, Bagheri Yazdi SA, Faghihzadeh S, Kamali K, Faghihzadeh E, Hajebi A, et al. A Survey on Mental Health Status of Adult Population Aged 15 and above in the Province of Semnan, Iran. *Arch Iran Med.* 2017;**20**(11 Suppl. 1):S103-6. [PubMed: 29481141].
- Noorbala AA, Bagheri Yazdi SA, Faghihzadeh S, Kamali K, Faghihzadeh E, Hajebi A, et al. Trends of Mental Health Status in Iranian Population Aged 15 and above between 1999 and 2015. *Arch Iran Med.* 2017;**20**(11 Suppl. 1):S2-6. [PubMed: 29481116].
- Li W, Zhang L, Luo X, Liu B, Liu Z, Lin F, et al. A qualitative study to explore views of patients', carers' and mental health professionals' to inform cultural adaptation of CBT for psychosis (CBTp) in China. *BMC Psychiatry.* 2017;**17**(1):131. doi: 10.1186/s12888-017-1290-6. [PubMed: 28390407]. [PubMed Central: PMC5385068].
- Lopez-Pina JA, Sanchez-Meca J, Lopez-Lopez JA, Marin-Martinez F, Nunez-Nunez RM, Rosa-Alcazar AI, et al. The Yale-Brown Obsessive Compulsive Scale: A Reliability Generalization Meta-Analysis. *Assessment.* 2015;**22**(5):619-28. doi: 10.1177/107319114551954. [PubMed: 25268017].
- Casu G, Gremigni P, Sommaruga M. The Patient-Professional Interaction Questionnaire (PPIQ) to assess patient centered care from the patient's perspective. *Patient Educ Couns.* 2019;**102**(1):126-33. doi: 10.1016/j.pec.2018.08.006. [PubMed: 30098906].
- Muntingh AD, van der Feltz-Cornelis CM, van Marwijk HW, Spinhoven P, Penninx BW, van Balkom AJ. Is the Beck Anxiety Inventory a good tool to assess the severity of anxiety? A primary care study in the Netherlands Study of Depression and Anxiety (NESDA). *BMC Fam Pract.* 2011;**12**:66. doi: 10.1186/1471-2296-12-66. [PubMed: 21726443]. [PubMed Central: PMC3224107].
- Irmak MK. Schizophrenia or possession? *J Relig Health.* 2014;**53**(3):773-7. doi: 10.1007/s10943-012-9673-y. [PubMed: 23269538].
- Bass JK, Bolton PA, Murray LK. Do not forget culture when studying mental health. *Lancet.* 2007;**370**(9591):918-9. doi: 10.1016/S0140-6736(07)61426-3. [PubMed: 17869621].
- Esfahani SR, Motaghipour Y, Kamkari K, Zahireidin A, Janbozorgi M. Reliability and Validity of the Persian version of the Yale-Brown Obsessive-Compulsive scale (Y-BOCS). *Iran J Psychiat Clin Psychol.* 2012;**17**(4). Persian.
- Dadfar M, Kalibatseva Z. Psychometric Properties of the Persian Version of the Short Beck Depression Inventory with Iranian Psychiatric Outpatients. *Scientifica (Cairo).* 2016;**2016**:8196463. doi: 10.1155/2016/8196463. [PubMed: 27293979]. [PubMed Central: PMC4886104].
- Beck AT, Steer RA, Brown G. *Manual for the beck depression inventory-II.* San Antonio, USA: Psychological Corporation; 1996. doi:10.1037/t00742-000.
- Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol.* 1988;**56**(6):893-7. doi: 10.1037//0022-006x.56.6.893. [PubMed: 3204199].

18. Cole MW, Yarkoni T, Repovs G, Anticevic A, Braver TS. Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *J Neurosci*. 2012;**32**(26):8988–99. doi: [10.1523/JNEUROSCI.0536-12.2012](https://doi.org/10.1523/JNEUROSCI.0536-12.2012). [PubMed: [22745498](https://pubmed.ncbi.nlm.nih.gov/22745498/)]. [PubMed Central: [PMC3392686](https://pubmed.ncbi.nlm.nih.gov/PMC3392686/)].
19. World Health Organization. *Culture and mental health in Haiti: A literature review*. World Health Organization; 2010.
20. Perbal B. Neuroscience and psychological studies sustain the cognitive benefits of print reading. *J Cell Commun Signal*. 2017;**11**(1):1–4. doi: [10.1007/s12079-017-0379-5](https://doi.org/10.1007/s12079-017-0379-5). [PubMed: [28155112](https://pubmed.ncbi.nlm.nih.gov/28155112/)]. [PubMed Central: [PMC5362581](https://pubmed.ncbi.nlm.nih.gov/PMC5362581/)].
21. Mahmoodi Z, Karimlou M, Sajjadi H, Dejman M, Vameghi M. Development of mother's lifestyle scale during pregnancy with an approach to social determinants of health. *Glob J Health Sci*. 2013;**5**(3):208–19. doi: [10.5539/gjhs.v5n3p208](https://doi.org/10.5539/gjhs.v5n3p208). [PubMed: [23618491](https://pubmed.ncbi.nlm.nih.gov/23618491/)]. [PubMed Central: [PMC4776819](https://pubmed.ncbi.nlm.nih.gov/PMC4776819/)].
22. Larwin K, Harvey M. A demonstration of a systematic item-reduction approach using structural equation modeling. *Pract Assess Res Evaluation*. 2012;**17**(1):1–19. doi: [10.7275/onem-w659](https://doi.org/10.7275/onem-w659).
23. Eslami M, Heidania A, Heidarzadeh A. Designing and determining validity and reliability of the questionnaire for the effect of HBM on users of two methods of birth control with pills and condoms. *Oromieh Journal*. 2011;**21**(5):382–90. Persian.
24. Borsa JC, Damásio BF, Bandeira DR. [Adaptation and validation of psychological instruments across cultures: some considerations]. *Paideia (Ribeirão Preto)*. 2012;**22**(53):423–32. Portuguese. doi: [10.1590/s0103-863x2012000300014](https://doi.org/10.1590/s0103-863x2012000300014).
25. Martin MS, Colman I, Simpson AI, McKenzie K. Mental health screening tools in correctional institutions: a systematic review. *BMC Psychiatry*. 2013;**13**:275. doi: [10.1186/1471-244X-13-275](https://doi.org/10.1186/1471-244X-13-275). [PubMed: [24168162](https://pubmed.ncbi.nlm.nih.gov/24168162/)]. [PubMed Central: [PMC4231452](https://pubmed.ncbi.nlm.nih.gov/PMC4231452/)].
26. Carson D, Grubin D, Parsons S. Report on New Prison Health Reception Screening Arrangement: The Result of a Pilot Study in Ten Prisons. Newcastle: University of Newcastle. 2003.
27. Taghavi S. Validity and reliability of the general health questionnaire (ghq-28) in college students of shiraz university. *J Psychol*. 2002;**5**(4):381–98.