Original Article

World Health Organization Quality-of-Life Scale (WHOQOL-BREF): Analyses of Their Item Response Theory Properties Based on the Graded Responses Model

Shahrum Vahedi, PhD[1]

1. Department of Educational Psychology, Faculty of Education and Psychology, University of Tabriz, Tabriz, Iran

Corresponding author:
Shahrum Vahedi, PhD
Assistant professor of Educational Psychology, Faculty of Education and Psychology, University of Tabriz. 22 Bahman Ave., Tabriz, Iran.

Tel: 0411-3392090
E-mail: vahedi117@yahoo.com

Objective: This study has used Item Response Theory (IRT) to examine the psychometric properties of Health-Related Quality-of-Life.
Method: This investigation is a descriptive- analytic study. Subjects were 370 undergraduate students of nursing and midwifery who were selected from Tabriz University of Medical Sciences. All participants were asked to complete the Farsi version of WHOQOL-BREF. Samejima's graded response model was used for the analyses.
Results: The results revealed that the discrimination parameters for all items in the four scales were low to moderate. The threshold parameters showed adequate representation of the relevant traits from low to the mean trait level. With the exception of 15, 18, 24 and 26 items, all other items showed low item information function values, and thus relatively high reliability from low trait levels to moderate levels.
Conclusions: The results of this study indicate that although there was general support for the psychometric properties of the WHOQOL-BREF from an IRT perspective, this measure can be further improved. IRT analyses provided useful measurement information and demonstrated to be a better methodological approach for enhancing our knowledge of the functionality of WHOQOL-BREF.

Keywords: *Item response theory, Psychometrics, Quality of life*

Iran J Psychiatry 2010; 5:140-153

In recent years, quality of life instruments have been acknowledged as very important in the evaluation of health care (1). Health-Related Quality of Life (HRQOL) refers to individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad-ranging concept affected in a complex way by the individual's physical health, psychological state, level of independence, social relationships, and their relationships to salient features of their environment" (2).

There are many general instruments available to measure quality of life. The World Health Organization (WHO) has developed a quality of life instrument, the WHOQOL, which captures many subjective aspects of quality of life) 3-5(. The WHOQOL-BREF is one of the best known instruments that has been developed for cross-cultural comparisons of quality of life and is available in more than 40 languages. It has been adopted in the United State of America, Netherlands, Poland, Bangladesh, Thailand, India, Australia, Japan, Croatia, Zimbabwe and many other countries (6 ,7). During the development of the WHOQOL, it was emphasized that quality of life is a multidimensional concept (5).According to international standards of WHO, including forward and backward translations, and focus group discussions, Nedjat et al. has translated the WHOQOL into Persian (8). An abbreviated version of the WHOQOL-BREF that contains 26 items is applicable in clinical trials in which brief measures are needed, and also in epidemiological studies in which quality of life might be one of several outcome variables (9). The WHOQOL BREF covers four different domains of quality of life (10). The WHOQOL is under cross-cultural validation by the WHOQOL group (5).

To-date, the studies that have examined the psychometric properties of the WHOQOL-BREF in Iran have all used scores based on the traditional classical test theory (7). Besides the CTT, another approach for examining the psychometric properties of measures is Item Response Theory (10-13).

IRT is a useful tool for gaining insights that traditional techniques cannot provide. Also, it is useful in screening items for inclusion in new questionnaires, and for checking the validity of assumptions even in traditional tests. On the account of these purposes, it certainly deserves wide usage. The most exciting roles for IRT in quality of life(QoL) research, however, lie firstly in the standardization of different instruments so

Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

that QoL as assessed by disease- and treatment-specific instruments can be compared across different groups of patients ;and secondly, in the development of computer-administered adaptive testing. Both of these objectives require extremely large databases for the exploration of IRT models. Many of these aspects of IRT are of obvious relevance to QoL assessment. Most QoL items, however, permit responses at more than two levels and multi-category IRT is far less developed. In addition, as with all models, IRT makes particular assumptions about the structure of the data, but some of these assumptions may be questionable when applied to QoL scales. IRT is mainly of relevance when considering scales that aim to classify patients into levels of ability, for example activities of daily living (ADL) or other physical performance scales (14).IRT is a model-based measurement theory that aims to show the relationship between responses to items and the ability or trait that each item is supposed to be measuring (13).

Additionaly, in IRT, the responses to items are used to obtain continuous scaled estimates of the underlying trait, called theta . In most computer programs, the values have a mean of zero and a standard deviation of one. Two common item parameters produced by IRT are the item difficulty parameter (also called the threshold parameter) and the item discrimination parameter (or slope). The threshold parameter (b) indicates the point on the scale of the latent trait where a person has a 1.5 probability of responding positively to the item, while the item discrimination parameter (a) is the ability of an item to discriminate people at different levels of the underlying trait below and above the threshold parameter (15).

In IRT analysis, graphs of trace lines or curves are generated for each item, showing the probability of a positive response to the items as a function of the underlying trait. For an item with dichotomous responses or only two response options (such as ‘‘yes’’ and ‘‘no’’), the trace lines are called item characteristic curves (ICCs).

RT models also provide information functions for each item and for all items together. These are called item information function and test information function, respectively. The information function of an item indicates the reliability of an item at different points of the underlying trait, while the test information function provides the reliability of all the items together at different trait levels. IRT also provides the standard error (SE) of the test information function. As the SE of a test information function is the inverse of the test information function. The SE and the test information function can be viewed as indicators of the precision of the test at different trait levels (13).

It has been argued that IRT has more advantages than traditional classical test theory (CTT) in term of evaluating the psychometric properties of measures (11). Three advantages are of particular relevance to this study. Firstly, for a trait, CTT provides a single score, which is derived from the scores of the different items comprising the scale, while IRT provide trait scores at the item level. Secondly, CTT assumes and provides one reliability only (such as internal constancy) value and one SE value for all levels of the scores obtained in a measure. In contrast, IRT provides the reliability of each item at different levels of the underlying trait, controlling for the characteristics (e.g., difficulty) of the items in the scale. Thirdly, CTT psychometric properties, such as reliability, item-total correlation and SE are sample dependent, which means properties can vary across samples. However, within a linear transformation, IRT psychometric properties are assumed to be sample independent or group invariant. As IRT provides parameters at the item level, this approach would allow the identification of items that are functioning differently in terms of their ability to discriminate and also represent and reliably measure the traits at different levels of the underlying trait. This, in turn, it can facilitate the development and revision of the measures. Therefore, it can be argued that the use of IRT will provide not only more valuable data on the psychometrics of the scales and items of the WHOQOL-BREF, but also provides useful directions for their improvement. With IRT being given these advantages, the aim of the current study was to (a) demonstrate the method of graded response model item analysis by calibrating item parameters, scoring individuals, and obtaining information levels; (b) displaying and describing the functioning of high, moderate and low information items; (c) demonstrating the relationship between reliability and information and showing how prespecified reliability can be obtained through consideration of total scale information. Of course, the main objective of this study was to investigate psychometric properties of the WHOQOLBREF by the use of the Samejima's graded response model.

Materials and Method

Participants

Subjects were 370 undergraduate students of midwifery and nursing (48% female and 52%male) selected from Tabriz University of medical sciences. Regarding test administration, researchers first provided instructions on how to answer the questions. Then, participants completed the questionnaires on their own. After completing the questionnaires, participants handed them to the researchers directly.

Procedure and Instruments

All data were collected via self-report. To enhance accuracy, all participants were informed that their responses would remain confidential.

The WHOQOL-BREF is a 26-item instrument consisting of four domains: physical health (7 items), psychological health (6 items), social relationships (3 items), and environmental health (8 items); it also contains QOL and general health items. Each individual item of the WHOQOL-BREF is scored from 1 to 5 on a response scale, which is stipulated as a five-

Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

point ordinal scale. The scores are then transformed linearly to a 0–100-scale (16,17). The physical health domain includes items on mobility, daily activities, functional capacity, energy, pain, and sleep. The psychological domain measures include self-image, negative thoughts, positive attitudes, self-esteem, mentality, learning ability, memory concentration, religion, and the mental status. The social relationships domain contains questions on personal relationships, social support, and sex life. The environmental health domain covers issues related to financial resources, safety, health and social services, living physical environment, opportunities to acquire new skills and knowledge, recreation, general environment (noise, air pollution, etc.), and transportation (7).

Statistical procedures

Because the WHOQOL-BREF has a polytomous response format with the response options graded, an IRT model appropriate for this item response format is Samejima's (1969) graded responses model (GRM). The GRM conceptualizes an item in terms of a series of k_1 or mi response dichotomies, where k is the number of response options. Thus, if there are four response options, there will be three response dichotomies; namely, the first category versus all other categories, the first and second response categories versus the third and fourth response categories, and first three response categories versus the fourth category. The trace lines reflecting these comparisons, are referred to as operator characteristic curves (OCCs), which represent the probability of an examinees' raw item response falling in or above a given category threshold conditional on the trait level (θ).

Each OCC provides the location of the appropriate threshold parameters (b), which is the trait (θ) level where there is a 1.5 probability of endorsing the relevant response option or higher response options. The number of β's for an item will correspond to the number of response dichotomies. In the GRM, the discrimination parameter (a) for all response options of an item is constrained to be equal. This constraint is not imposed across items. Thus, each item will have its own single discrimination parameter. Once the threshold and discrimination parameters for the different response dichotomies of an item are known, the probability of response to each response option in the item as a function of the underlying trait can be generated. The resulting trace lines are called category response curves (CRCs). The CRC for the first response option will be a monotonically decreasing logistic function, while the CRC for the last response option will be a monotonically increasing logistic function. The CRCs for the other response options will all be nonmonotonic logistic functions.

Therefore, this study used Samejima's (18) GRM. All analyses were conducted with Multilog 7.0.3 (19). Unidimensionality was tested using Cronbach's alpha, Confirmatory Factor Analysis and Exploratory Factor Analysis procedures. In relation to the exploratory procedures, Parallel Analysis (20), followed by principal component analysis (PCA) were conducted for the WHOQOL-BREF Questionnaire. The PA procedure is known to provide more accurate results for the number of factors to be extracted than the eigenvalue greater than 1 (or K1) rule or the scree test (21). PA was conducted using the software of Monte Carlo PA provided by Watkins. The ratio of the eigenvalues of the first and second unrotated components from the real data sets can also be used for evaluating unidimensionality, with high ratios being indicative of unidimensionality.

The CFA approach involved testing the fit of 1-factor models for the WHOQOL-BREF scale. The 1-factor CFA models were tested using Amos software (22). Maximum Likelihood Estimates was used for the analyses. Fit was examined using the incremental fit index (IFI) and the comparative fit index (CFI).

Results

Demographics of the current sample

Data were collected from 370 undergraduate students of nursing and midwifery(51.4%) males and 170 females (45.9%).

Subjects were randomly selected from Tabriz University of medical sciences.

The average age of the participants was 21.55±2.39 year. Approximately 16.8%, (n=52) of the subjects were living in rural areas, but the majority of them were living in cities (83.2%, n=190).

Checking model assumptions

Two primary assumptions of IRT are unidimensionality of the scale, and local independence, which posits that when the respondent trait levels are controlled for, the items on the scale are independent from one another (11, 21).

This unidimensionality is important because the basic Samejima's model assumes unidimensionality. In order to ensure this, we conducted a confirmatory factor

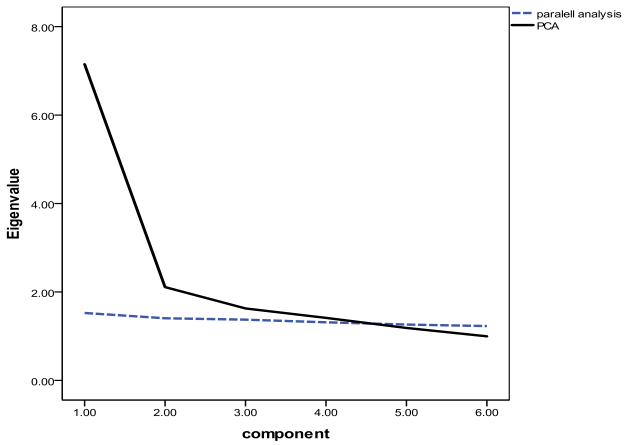


Figure 1. Scree plot of the WHOQOL-BREF at baseline and exit with randomly generated scree (parallel analysis).

Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

analysis (CFA) on the 26 WHOQOL-BREF items. A confirmatory factor analysis of the four a priori domain scales of the WHOQOL-BREF improved fit over the one-factor model ($P < 0.001$). Overall, model fit was good: Bollen's incremental fit index (IFI) =0.940, comparative fit index (CFI) =0.939, Bentler-Bonett normed fit index (NFI) =0.933 and the factor loadings for each of the subscales ranged from 0.48 to 1.33.

In addition, Fig. 1 indicates, the WHOQOL was one-dimensional at baseline and at exit in that the observed first component was always smaller than that of generated from random data. The explorative factor analysis, parallel factor, in conjunction with the confirmatory factor analysis, meet the assumption of a general WHOQOL-BREF dimension underlying each scale.

IRT analyses of the personal scale of the WHOQOL BREF

Table 1 demonstrates five parameter estimates/items for all the items of the subscales of the WHOQOL BREF –one slope (a) ,and four threshold separating the five response categories (b0, b1, b2, and b3).

In IRT, these values of a can range from 0 to around 3. They represent how quickly an item's scores change as a function of changes in the latent trait. Like factor loadings in a CFA, they capture how closely an item represents the latent trait being measured.

As shown in this table, with the exception of item 15, values of α for all items were large, although some are higher than others. Based on Baker's (23) guidelines, fifteen items have moderate discrimination (3,4,7,11-14,16,20-26), four items have high discrimination (6,9,18 ,19), two items have very high discrimination (2,17), and three items have perfect discrimination (5,8 ,10).

To get a sense of what the discrimination values mean, we can view the category response curves (CRC), which display how the score probabilities vary as a function of the latent curiosity trait.

For the sake of illustration, Fig. 2 presents the CRCs for all items that differ in their discrimination values. The CRCs depict the probability that someone will respond to the item with a 1, 2, 3, 4, or 5. The probability level is along the Y-axis, and the level of trait quality of life, expressed as a standard normal distribution (M=0, SD=1) is along the X-axis.

The different discrimination values are evident in the peaks and overlaps of the probability curves.

For the items 2, 5, 8, 9, 17 and 10, which have a higher discrimination values, the response probabilities have higher peaks and have relatively less overlap. Each scale value (1 through 5) has a region where its absolute probability is at least 50%, indicating that there are trait levels where that response is more probable than the other four options combined. For other items, in contrast, the response probabilities are flatter and overlap more significantly. For the middle three options (2-4), the response curves are never higher than 0.50.

Difficulty Thresholds

Each item has five possible responses, so there are four response thresholds depicted as b1, b2, b3, and b4. In IRT, these thresholds represent the trait levels at which someone has a 50% chance of scoring at or above a scale response.

These thresholds give a good deal of information about each item. For example, we will use the thresholds for item 1. The value for b1, the threshold between a response of 1 and a response of 2, is -3.16. This means that someone with a trait score of -3.16 , a pretty low level, has at least a 50% chance of responding to the item with a 2, 3, 4, or 5. Conversely stated, a response of 1 is the most likely response for people with a quality of life less than -3.16. The value for b2, the threshold between the scores of 2 and 3, is -2.02. Therefore, those with a quality of life score of -2.02 have at least a 50% chance of responding with a 3, 4, or 5. On the whole, these thresholds reveal that item 1 is very easy. The last threshold, b4, is 0.84, which is only around .12 SD above the mean. This means that a student with a quality of life greater than .84 has at least a 50% chance of responding with a 5, the highest possible scale score. The difficulty thresholds are related to the frequency with which people chose different response options. Item 25, for example, has very high b4 threshold, and relatively few people responded with a 5 to the item. Conversely, item 3 has very low b1 threshold, and relatively few people responded with a 1 to the item.

For a self-report scale that measures individual differences, it is desirable for the items to offer information about a broad range of the trait. For this reason, some items should be ‘‘easier" and others should be ‘‘harder." An example of a relatively easy item is item 3-its lowest threshold is quite low, and only 1.1% of the sample responded with a 1 to it. Only people who are very low in quality of life will respond with low scores to this item. A relatively harder item is item 15: its highest threshold is 4.66, so only people who are very high in quality of life will respond with a five. Nevertheless, there is not much between-item variation in the difficulty ranges. Each item covers a good range of the trait, but the items as a group tend to be centered at the trait's midpoint.

Moreover, Table 1 shows the threshold parameters (b1, b2, b3 and b4) for all the personal items. As it is shown in this table, the b parameter values made noticeable increases in the level of the latent trait at each subsequent response dichotomy. Also, the trait values for b1, b2 and b3 were somewhat evenly spaced with all values below the mean trait level. For all items, the trait values for b4 were only slightly above the mean trait level.

Table 3 demonstrates the item information function values of the twenty six items in the WHOQOL BREF scale.

Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

Table 1 IRT parameter estimates for the scales of the 26-Item WHOQOL BREF

item	α	ß1	ß 2	ß 3	ß 4
physical health					
3 Pain and discomfort	0.99(0.15	-4.20 (0.75)	-2.73 (0.44)	-1.10 (0.23)	0.51 (0.18)
4 medical treatment	0.91 (0.19)	-5.29 (1.27)	-4.48 (0.97)	-3.03 (0.62)	-1.40 (0.30)
10 Energy	2.15 (0.21)	-2.96 (0.41)	-1.62 (0.16)	-0.32 (0.09)	1.10 (0.11)
15 discomfort	0.48 (0.12)	-4.71 (1.30)	-1.37 0.49)	1.36 (0.46)	4.67 (1.42)
16 Sleep	0.90 (0.13)	-3.11 (0.52)	-1.71 (0.31)	0.22 (0.17)	2.40 (0.39)
17ability to perform daily living activities	1.76 (0.19)	-2.60 (0.31)	-1.44 (0.17)	0.12 (0.10)	1.76 (0.19)
18 capacity for work	1.32 (0.17)	-3.12 (0.44)	-1.65 (0.23)	0.08 (0.12)	1.73 (0.23)
Psychological (domaine2)					
5 Positive feelings	2.05 (0.20)	-2.09 (0.22)	-1.35 (0.14)	0.19 (0.09)	1.36 (0.13)
6 Self-esteem	1.69 (0.18)	-2.50 (0.32)	-1.62 (0.18)	-0.34 (0.11)	0.70 (0.12)
7Thinking, learning, memory and concentration	1.04 (0.14)	-3.36 (0.53)	-1.74 (0.28)	0.69 (0.17)	2.98 (0.47)
11Bodily image and appearance	1.26 (0.15)	-2.78 (0.38)	-1.66 (0.23)	-0.45 (0.14)	0.88 (0.16)
19 satisfy with you	1.53 (0.17)	-2.60 (0.34)	-1.53 (0.19)	0.63 (0.12)	2.22 (0.25)
26 Negative feelings	1.15 (0.14)	-2.56 (0.36)	-1.13 (0.20)	0.34 (0.14)	1.71 (0.24)
social relationships					
20 Personal relationships	1.31 (0.15)	-3.33 (0.46)	-1.73 (0.22)	0.25 (0.12)	2.12 (0.26)
21 Social support	1.06 (0.15)	-1.43 (0.28)	-0.44 (0.17)	0.87 (0.19)	2.47 (0.38)
22 Sexual activity	0.79 (0.13)	-3.18 (0.60)	-1.27 (0.30)	1.26 (0.28)	3.66 (0.67)
environmental health					
8Freedom, physical safety and security	2.26 (0.21)	-2.36 (0.26)	-1.37 (0.13)	-0.22 (0.09)	1.35 (0.12)
9 Physical environment	1.68 (0.18)	-3.17 (0.48)	-2.17 (0.26)	-0.29 (0.11)	1.52 (0.18)
12 Financial resources	1.02 (0.15)	-2.39 (0.39)	-1.39 (0.24)	0.26 (0.16)	2.08 (0.33)
13Opportunities for acquiring new information and skills	1.17 (0.15)	-3.70 (0.58)	-1.34 (0.22)	0.65 (0.15)	2.69 (0.38)
14Participation in and opportunities for recreation/leisure	0.87 (0.14)	-2.94 (0.51)	-0.41 (0.19)	1.51 (0.29)	3.32 (0.59)
23 Home environment	1.20 (0.15)	-2.20 (0.31)	-1.38 (0.21)	0.40 (0.13)	2.12 (0.28)
24Health and social care: accessibility and quality	0.92 (0.14)	-3.00 (0.51)	-1.40 (0.27)	0.45 (0.17)	2.53 (0.41)
25 Transport	0.82 (0.13)	-2.47 (0.45)	-1.13 (0.26)	1.17 (0.26)	3.39 (0.61)
Overall Quality of Life and General Health					
1 Overall Quality of Life	1.84 (0.19)	-3.16 (0.48)	-2.02 (0.22)	-0.55 (0.10)	0.84 (0.12)
2 General Health	1.05 (0.15)	-1.42 (0.28)	-0.43 (0.17)	0.88 (0.19)	2.48 (0.38)

α - is the discrimination parameters
β - is the difficulty parameter

As shown in this table, with the exception of item on 10, the information values of all the items were quite low at all trait levels. The item 10 had relatively high information values from trait values _3.0 to 2.0.
Figure 4 displays the test information function for the 4 quality of life scales. As can be seen in these plots, with the exception of items 10, most of the curves are completely low. This indicates that the overall degree of measurement precision for these items is also relatively low. Specifically, the items are less precise for measuring individuals with theta levels falling above 1.00 (e.g., 3 item) and below-1(e.g., 21 item).

Total information curve

The Test Information functions (TIF) provides information on the reliability of the WHOQOL-BREF across the range of latent–trait scores and is computed via a combination of location of the item and discrimination parameters. As noted by Neal and colleagues (24), when the total information curve is generally peaked, it indicates the highest degree of reliability the scale has at that level of the underlying latent trait score. As shown in Fig.3, the total information curve indicates that the WHOQOL BREF has relatively moderate reliability. Note that Test information functions for the scale are relatively flat in the low range of θ continuum. It does not have distinct peaks, but generally provides most information for θ levels between zero and -2 (measured in standard deviations). Their accuracy sharply decreases as θ increases over 1.
Test information functions for WHOQOL BREF scale appears to be the most evenly spread out across whole range of θ continuum. Indeed, this scale measures respondents with different levels of quality of life, from low to moderate, with almost equal precision. However, precision of the scale decreases sharply, while standard error of measurement increases sharply when respondents with high quality of life are measured .

Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

Additional reliability and validation analyses of the WHOQOL BREF.

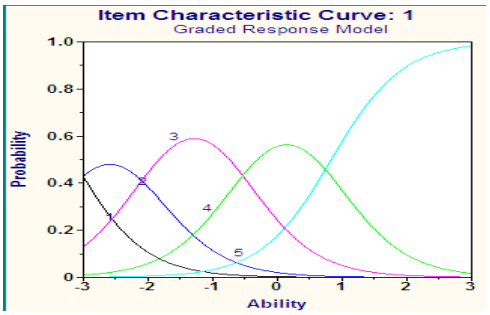
Additional analyses focused on examining the reliability and validity of the refined WHOQOL BREF via classical measurement techniques. For these analyses, Classical Test Theory (CTT) statistics were conducted for the original 26-item WHOQOL BREF.

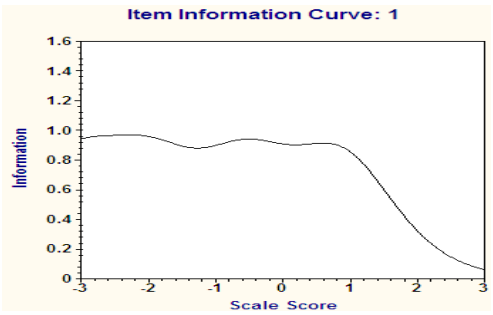
The Cronbach's alpha values for physical health, psychological health, social relationships and environmental health were 0.65, 0.77, 0.52 and 0.79, respectively. The mean item-to-total correlations were 0.76, 0.73, 0.62 and 0.78 for physical health, psychological health, social relationships and environmental health, respectively. For each WHOQOL-BREF , the factor analysis resulted in only one factor. Taken together, these findings support the unidimensionality of the four scales and the local independence of the items in each scale.

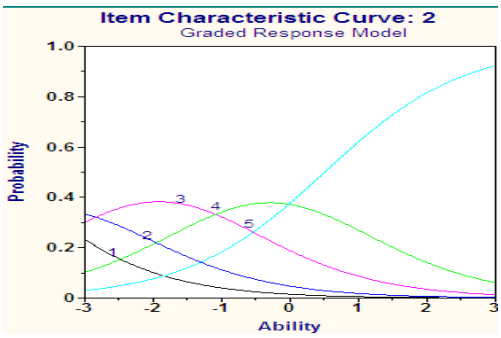
Table 2 Information for the items in the four scales of the spiritual well-being questionnaire

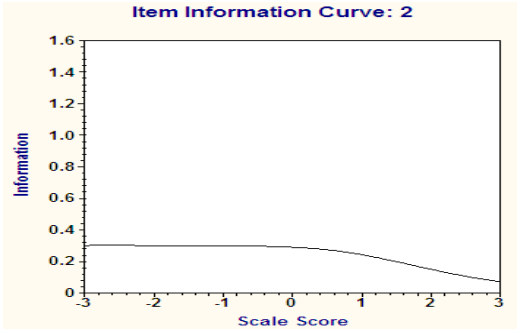
Item	Estimated trait						
	-3	-2	-1	0	1	2	3
physical health							
3 Pain and discomfort	0.30	0.30	0.30	0.29	0.24	0.15	0.07
4 medical treatment	0.26	0.25	0.14	0.21	0.08	0.03	0.02
10 Energy	1.21	1.21	1.19	1.20	1.22	0.51	0.08
15 discomfort	0.07	0.07	0.07	0.07	0.07	0.07	0.07
16 Sleep	0.23	0.25	0.25	0.24	0.24	0.23	0.20
17 ability to perform daily living activities	0.70	0.89	0.83	0.85	0.78	0.75	0.28
18 capacity for work	0.49	0.52	0.50	0.51	0.50	0.44	0.23
Psychological (domaine2)							
5 Positive feelings	0.49	1.21	1.09	1.12	1.16	0.70	0.14
6 Self-esteem	0.61	0.86	0.82	0.85	0.70	0.26	0.06
7Thinking, learning, memory and concentration	0.31	0.32	0.30	0.30	0.30	0.30	0.30
11Bodily image and appearance	0.42	0.49	0.49	0.48	0.43	0.25	0.10
19 satisfy with you	0.55	0.69	0.59	0.57	0.65	0.64	0.42
26 Negative feelings	0.33	0.39	0.40	0.40	0.40	0.34	0.20
social relationships							
20 -Personal relationships	0.49	0.50	0.47	0.48	0.47	0.47	0.32
21- Social support	0.15	0.27	0.34	0.35	0.35	0.33	0.27
22 -Sexual activity	0.18	0.19	0.19	0.18	0.18	0.18	0.18
environmental health							
8 Freedom, physical safety and security	0.80	1.43	1.31	1.38	1.23	0.78	0.12
9 Physical environment	0.81	0.78	0.68	0.74	0.71	0.61	0.20
12 Financial resources	0.25	0.31	0.32	0.32	0.31	0.29	0.21
13 Opportunities for acquiring new information and skills	0.36	0.36	0.39	0.38	0.39	0.38	0.34
14 Participation in and opportunities for recreation/leisure	0.20	0.21	0.22	0.23	0.23	0.23	0.22
23 Home environment	0.30	0.42	0.43	0.42	0.42	0.40	0.28
24 Health and social care: accessibility and quality	0.24	0.26	0.26	0.26	0.25	0.25	0.21
25 Transport	0.17	0.20	0.21	0.20	0.20	0.20	0.19
Overall Quality of Life and General Health							
1 Overall Quality of Life	0.14	0.26	0.33	0.35	0.34	0.33	0.27
2 General Health	0.94	0.96	0.90	0.91	0.85	0.32	0.06

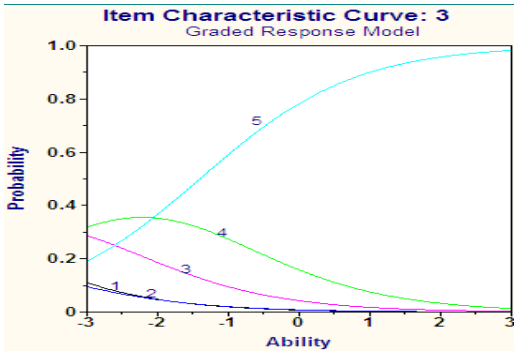
Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

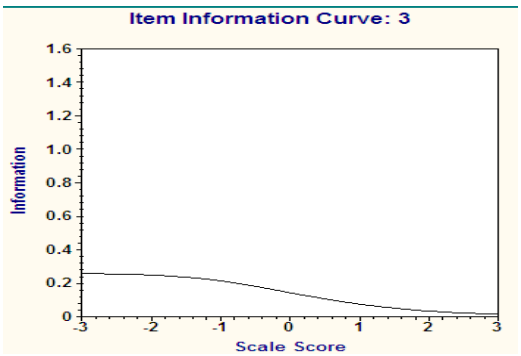


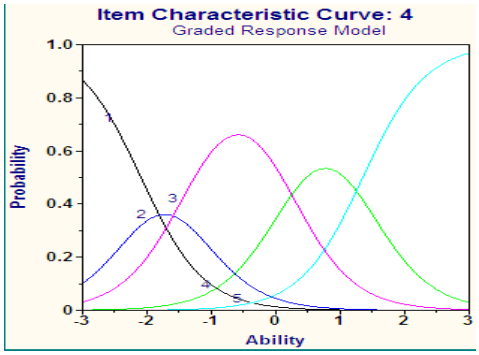


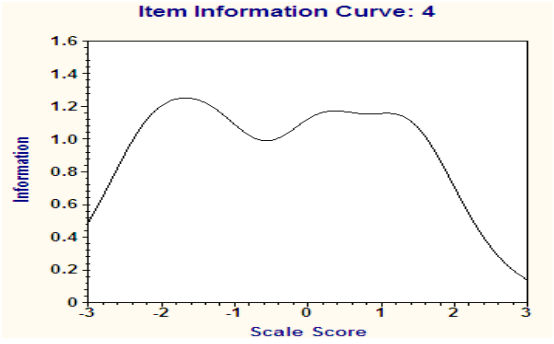




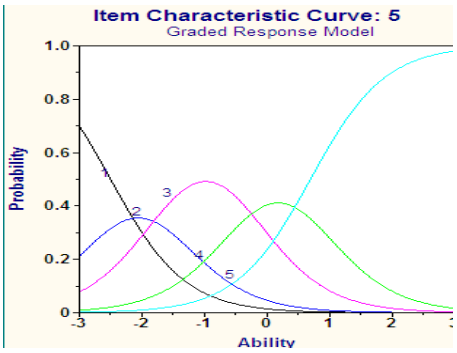


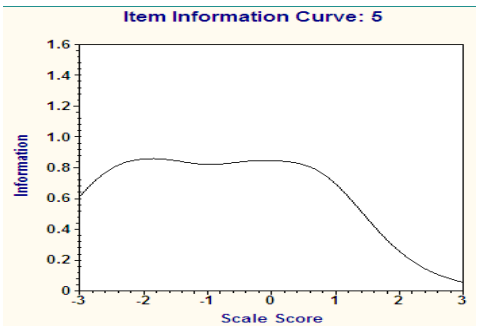


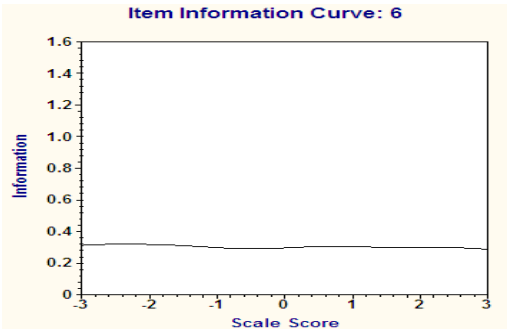


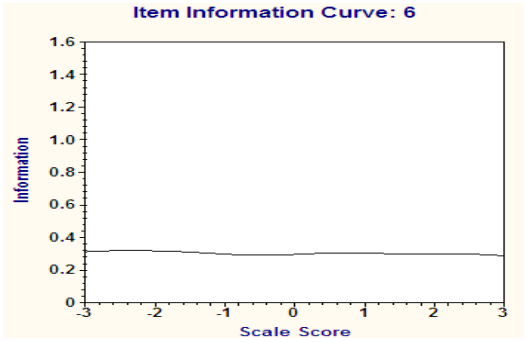


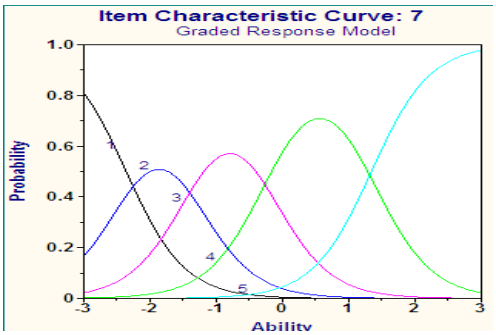
Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

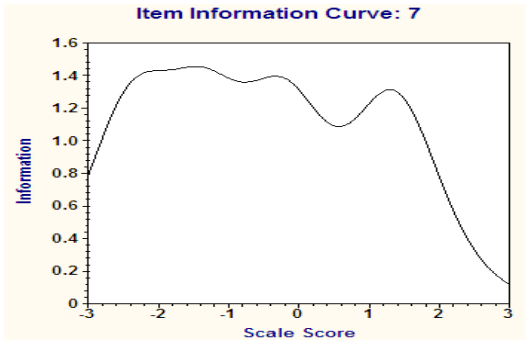


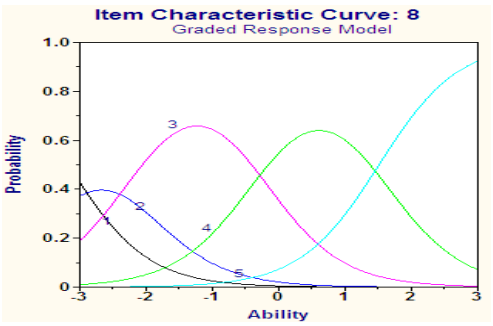


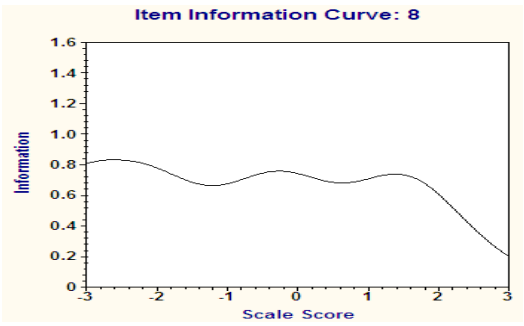


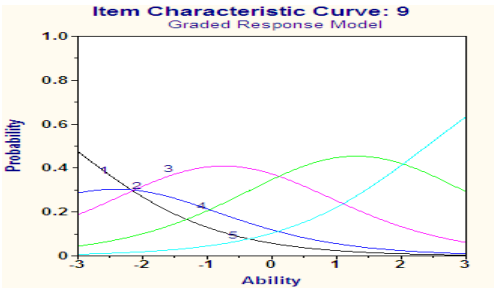


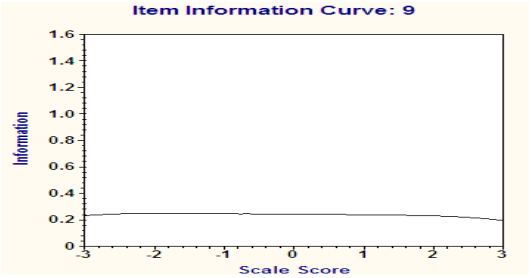




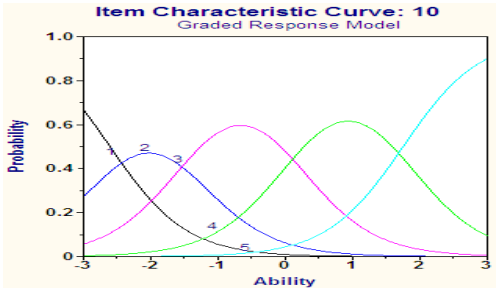


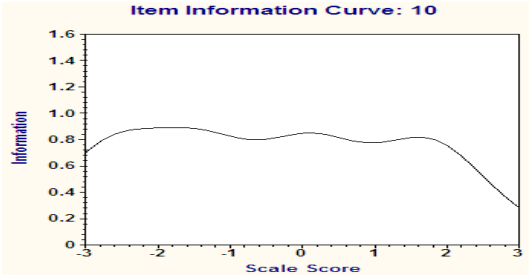


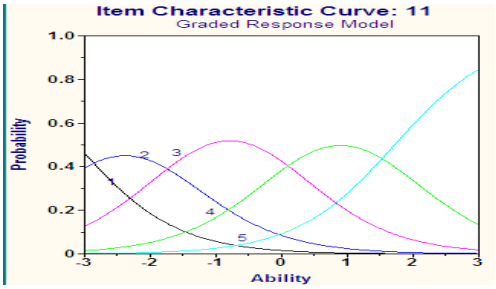


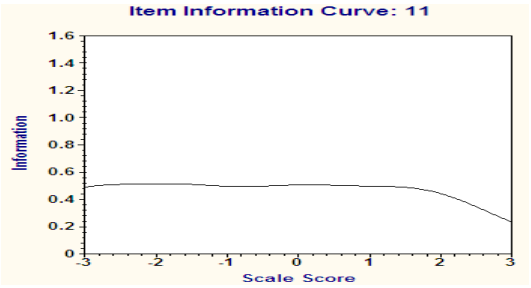


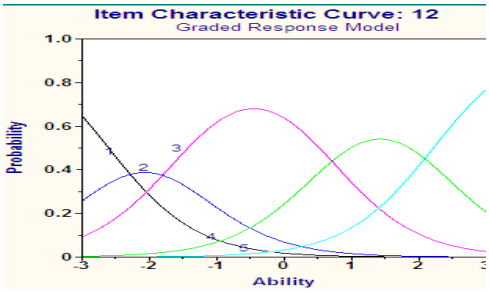
Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

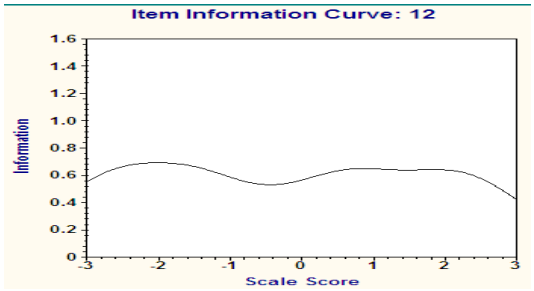


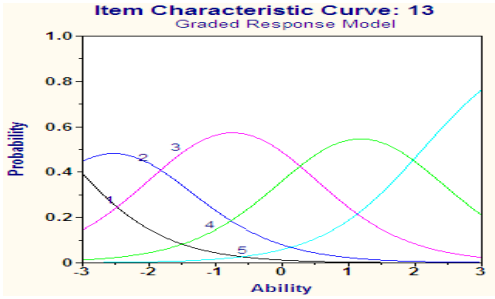


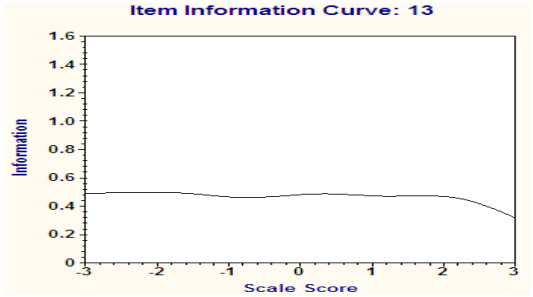


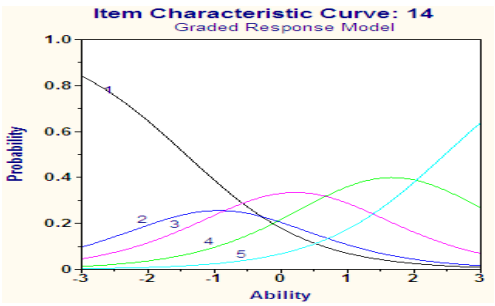


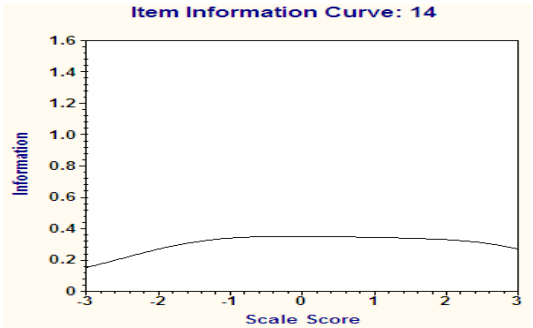




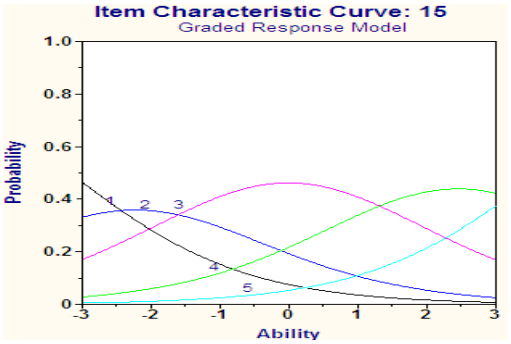


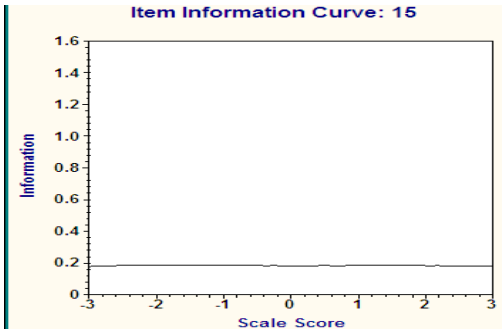


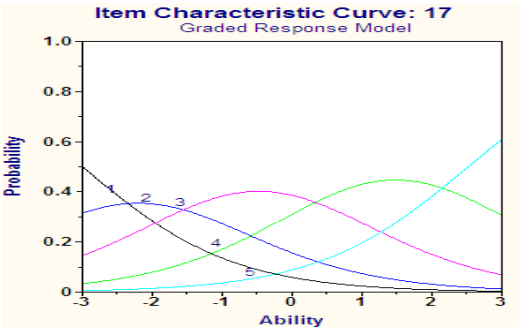


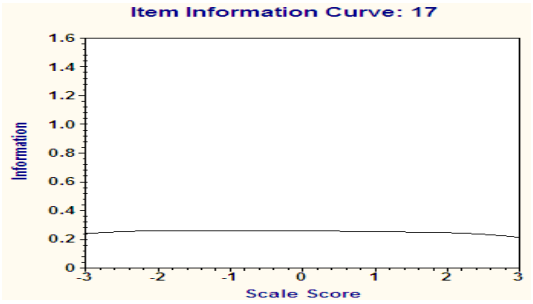


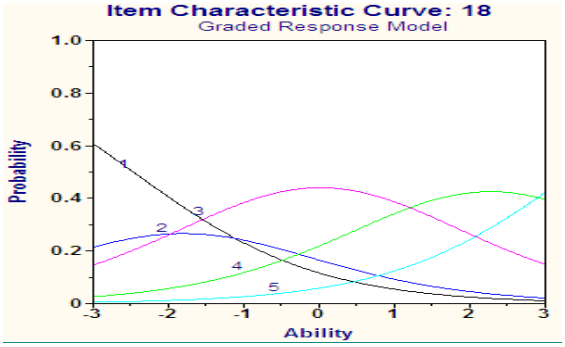
Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

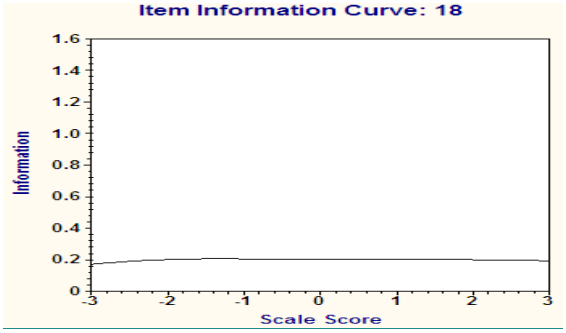


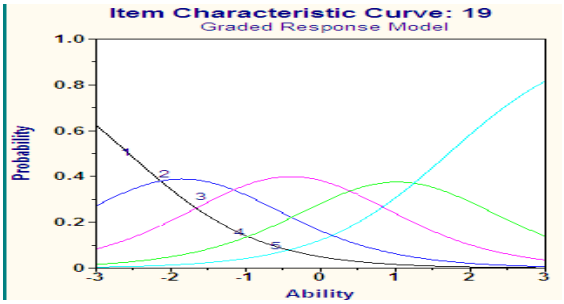


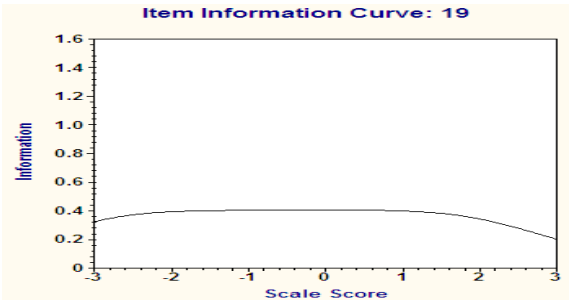




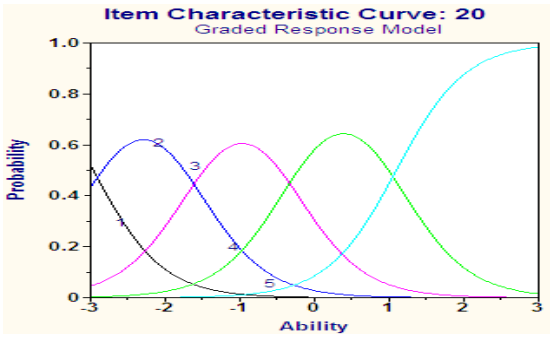


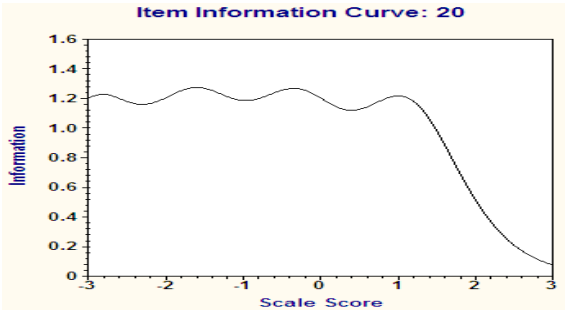


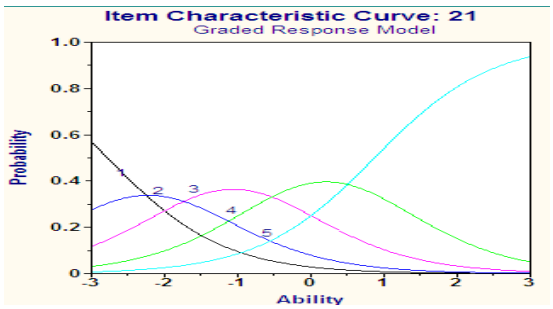


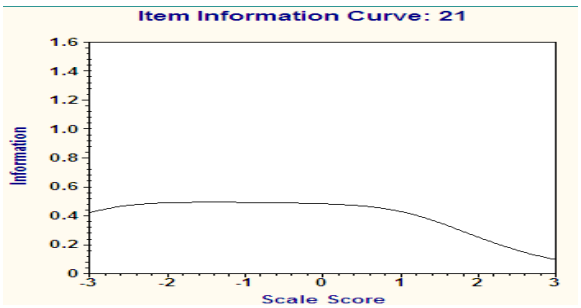


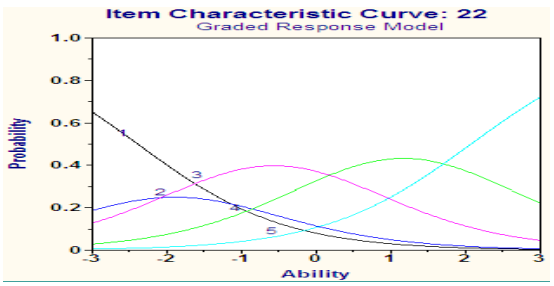
Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

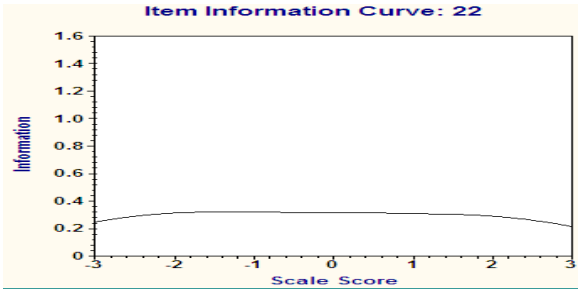


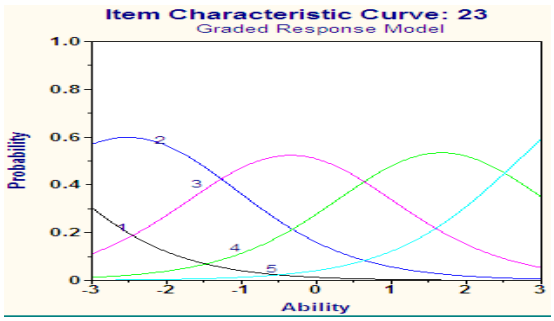


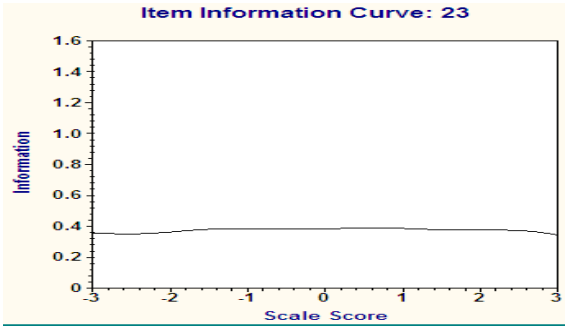


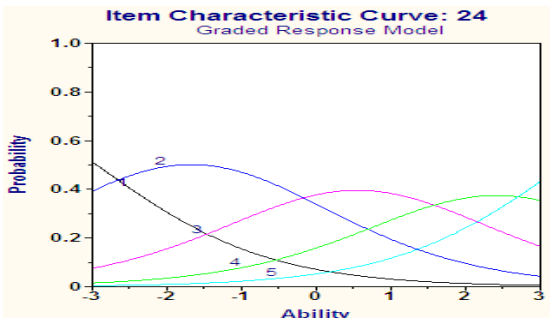


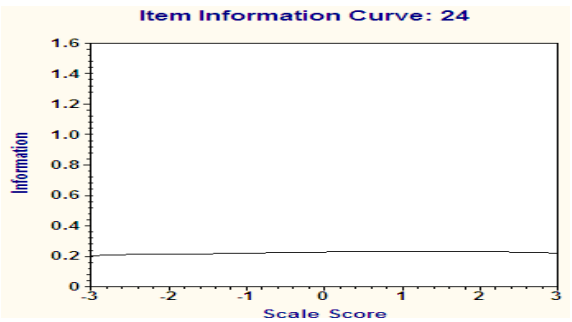




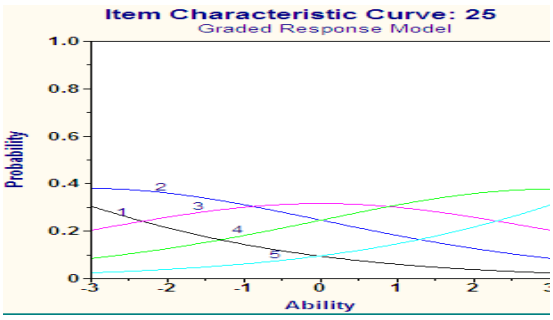


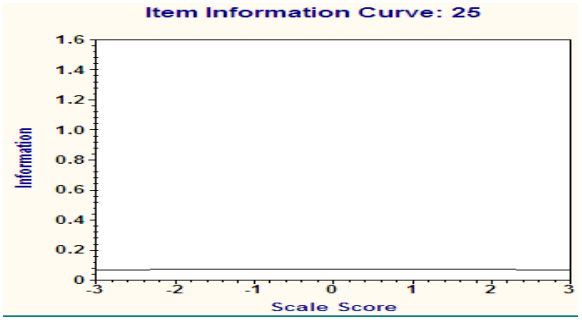






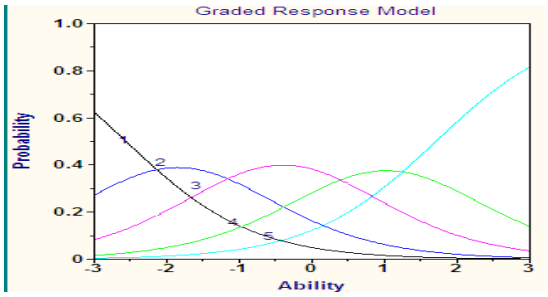
Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

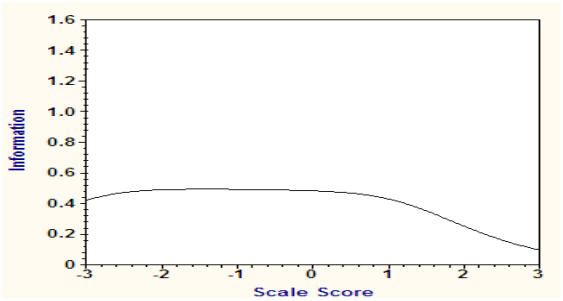




item characteristic Curve:27

item characteristic Curve:28





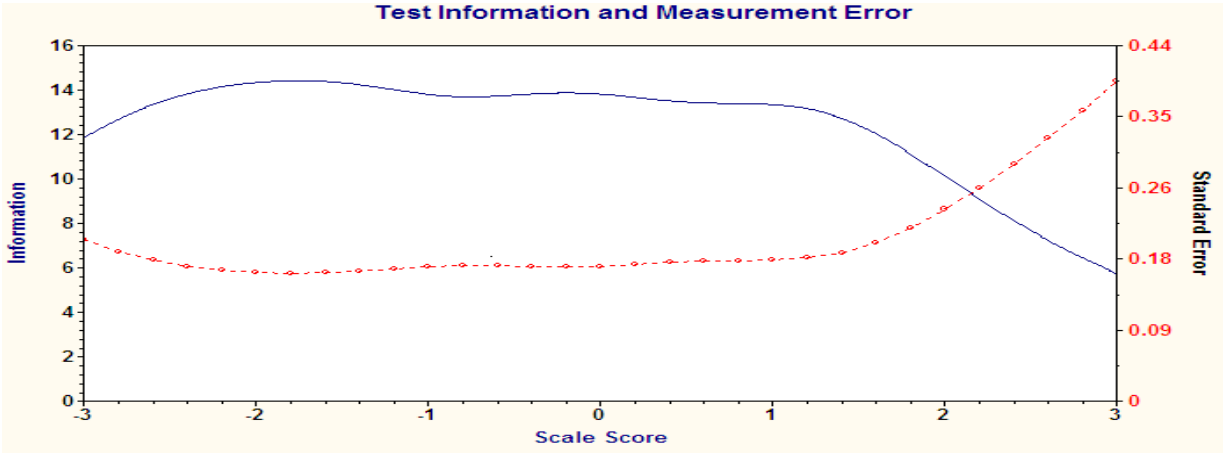


Fig. 2. Illustration of category response and item response functions, and test information function (continuous lines) and standard error curves for the four scales of the quality of life.

Discussion

The main purpose of this study was to evaluate psychometric properties of WHOQOL-BREF, in order to determine how useful the instrument is for research, training and development, self-evaluation, self-improvement, promotion and compensation purposes. Current psychometric evaluation of the WHOQOL-BREF has been limited to Classical Test Theory (CTT) techniques. In recent years, IRT methods have been used to develop measurement tools for health status assessment, for example, to construct instruments, score scales, or to validate tests. These applications have focused on the measurement component of IRT. Item Response Theory was used to achieve this end. Methods from IRT, however, offer several advantages in comparison to methods from CTT such as providing information on the performance of individual scale items and determining an instrument's reliability across the underlying trait that is purported to measure. Recently, these methods have been used to refine measures of health problems (24,25). Data from the WHOQOL-BREF were analyzed using F. Samejima's graded response model. Results show that the model fits well and is suitable for the analysis of the questionnaire.

The findings for the discrimination parameters of the items showed that with the exception of item 15, 18, 24 and 26, the remainder of all items can discriminate adequately well for low levels of subscales of

Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

WHOQOL-BREF, although this ability varies across items. More specifically, the items on 2, 5, 8, 6, 9, 17 and10 can discriminate equally well and better than the other items (3, 4, 7,11-14, 16, 18- 26). This finding is consistent with those obtained by Lin & Yao (27). One possible explanation for this finding is that questions with reverse coding (e.g. item 3, 4 and26) have poorer discriminatory power than un-recoded items. Indeed, reverse coded items provide different validity even if they have been recorded. The effect of wording has been examined by several studies. Zumbo et al. used item-total correlation and IRT models to test Roskam's conjecture and found inconsistent effects of wording and item. Another study conducted by Herche and Engelland also found that the use of reversed-polarity items led to some problems. A mixed use of positively and negatively worded items in the same scale can adversely affect measurement consistency. The reason for this is the degradation of scale dimensionality resulting from bias. Therefore, they recommended against using negatively worded items (27).

Furthermore, one explanation for this finding is that some items (3, 4, 6, 9, 11- 14 , 26) were so specific that they were more suitable for assessing a particular subgroup rather than a general population, and as a consequence these items also provided little information when assessing the QOL of a relatively healthy population in this case. In other words, general questions (1, 2 ,17) overall performed better than object-specific items.

Overall, IRT analysis shows that WHOQOL-BREF measure appears to be a moderately reliable instrument, better suited for identifying moderate quality of life levels. There are some caveats that the reader should be aware of when interpreting these results. First, the sample size of this study was relatively small (n ▯. 370). Large samples of examinees are required to accurately estimate the IRT item parameters. This results from either improved estimation of the θs or improved estimation of the shape of the θ distribution. In addition, increasing the number of examinees can somewhat improve the estimation of θ through improved estimation of the item parameters (28). In other words, for polychromous models, the number of examinees per category can make a difference in the accuracy of the category parameter estimation, and the thresholds are more accurately estimated for middle categories than for extreme categories. The category parameters are also estimated better when examinees are distributed more evenly across categories (28).

Second, the accuracy of estimating θ increases with the number of items. Better-quality items will be more discriminating and more useful for estimating θ. The accuracy of the item parameter estimates obviously has some impact on θ estimation as well, so increasing the examinee sample size and thus increasing the item parameter accuracy can increase the precision of θ estimation. Therefore, WHOQOL-BREF scales could be improved by removing some of the redundant items with low discrimination power (e.g. 24) and by adding new, and more "difficult" items located in the upper part of the quality of life ability (θ) continuum in university population. Thus, the items with low reliability may need to be revised to improve their reliability.

Indeed, the overall results of this study are supportive of its psychometric properties, so has been the result of previous study (26). This study also was in accordance with the work of Lin & Yao (27) who reported that a more precise and valid measure of the QOL can be constructed using the IRT approach. It is important not only to develop a comprehensive but also short and easily administered QOL instrument, which will have a significant impact for clinical or research purposes.

In conclusion, this study has suggested that the use of IRT procedures can provide valuable additional psychometric information. It is well documented that good CTT- based psychometric properties for a measure do not necessarily mean that it would have good IRT-based psychometric properties. This has to be demonstrated using IRT procedures. This study has also demonstrated how IRT can be used to revise the existing measures. It is hoped that this study could be able to show the value of using IRT to evaluate the psychometric properties of measures, test development and revision, and that it would encourage other researchers to use IRT approaches for similar purposes. Researchers should keep in mind that IRT analyses are affected by response tendencies and social desirability. Therefore, additional research is needed on the WHOQOL-BREF, especially efforts designed to increase measurement precision at the high-end of the scale, as well as studies of Differential Item Functioning (DIF) and Computer Adaptive Testing (CAT).

There were some limitations associated with the current study. The reliance on cross-sectional assessment strategies and college student samples raise questions about the generalizability of the findings. We are hopeful that the ongoing result will be examined in other samples with different age groups and different socio-cultural backgrounds so it will not be limited to university students.

Acknowledgement

We are tremendously grateful to all the participants of this study.

References

1. WHOQOL Group.study protocol for the World Health Organization project to develop a Quality of life assessment instrument (WHOQOL). Qual Life Res1993; 2: 153–159.
2. Carr A, Higginson I, Robinson PG. Quality of Life,Volume 13. BMJ Books; 2003.

Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)

3. WHOQOL Group. Development of the WHOQOL: Rationale and current status. Int J Ment Health 1994; 23: 24–56.
4. The WHOQOL Group. The World Health Organization Quality of Life assessment (WHOQOL): Development and general psychometric properties. Soc Sci Med 1998; 46: 1569–1585.
5. Noerholm V, Groenvold M, Watt T, Bjorner JB, Rasmussen NA, Bech P. Quality of life in the Danish general population – normative data and validity of WHOQOL-BREF using Rasch and item response theory models. Qual Life Res 2004; 13: 531–540.
6. World Health Organization's Quality of Life group: Measuring Quality of Life; Development of the World Health Organization Quality of Life Instrument (WHOQOL); 1992.
7. Nejat S, Montazeri A, Holakouie Naieni K, Mohammad K, Majdzadeh SR. [The World Health Organization Quality of Life (WHOQOLBREF) questionnaire: Translation and validation study of the Iranian version]. Journal of School of public Health & Institute of Public Health Research 2006; 4: 1-12.
8. Nørholm V, Bech P. The WHO Quality of Life (WHOQOL) Questionnaire: Danish validation study. Nord J Psychiatry 2001; 55: 229–235.
9. Brinbaum A. Some latent trait models and their use in inferring an examinee's ability. In: Lord FM , Novick MR,eds. Statistical theories of mental test scores. Reading, MA: Addison-Wesley;1968.
10. The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF Quality of Life Assessment. Psychol Med 1998; 28: 551–558.
11. Embretson SE, Reise SP. Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum Associates;2000.
12. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago ,IL: University of Chicago Press;1960.
13. Gomez R, Fisher JW. Item response theory analysis of the spiritual well-being questionnaire. Pers Individ Dif 2005;38: 1107–1121
14. Fayers PM , Machin. DC. Quality of Life: The assessment, analysis and interpretation of patient-reported outcomes.chichester: John Wiley & Sons Ltd; 2007.
15. Steinberg L, Thissen D. Item response theory in personality research. In: Shrout PE , Fiske ST, eds, Personality research, methods, and theory: a festschrift honoring Donald W. Fiske. Hillside, NJ: Lawrence Erlbaum Associates;1995.
16. Skevington SM, Tucker C. Designing response scales for cross-cultural use in health care: Data from the development of the UK WHOQOL. Br J Med Psychol 1999; 72: 51–61.
17. Harper A, Power M. WHOQOL User manual. Edinburgh; 1999.
18. Samejima F. Estimation of latent trait ability using a response pattern of graded scores. Psychometrika Monograph Supplement 1969; 17:1-100.
19. Thissen D. MULTILOG: multiple category item analysis and test scoring using item response theory. Chicago:Scientific Software;1991.
20. Horn JL. A rationale and a test for the number of factors in factor analysis. Psychometrika1965; 30: 179–185.
21. O'Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test.Behav Res Methods Instrum Comput 2000; 32: 396–402.
22. Arbuckle J L. Amos user's guide version 18.0. Chicago: SmallWaters Corporation;2009.
23. Baker FB. The basics of item response theory. Portsmouth, NH: Heineman;1985.
24. Neal DJ, Corbin WR, Fromme K. Measurement of alcohol-related consequences among high school and college students: Application of item response models to the Rutgers alcohol problem index. Psychol Assess 2006;18:402-414.
25. Kahler CW, Strong DR., Hayaki J, Ramsey SE, Brown RA. An item response analysis of the Alcohol Dependence Scale in treatment-seeking alcoholics. J Stud Alcohol Drugs 2003; 64: 127−136.
26. Sijtsma K, Emons WH, Bouwmeester S, Nyklı´c˘ek I, Roorda LD. Nonparametric IRT analysis of Quality-of-Life Scales and its application to the World Health Organization Quality-of-Life Scale (WHOQOL-BREF). Qual Life Res2008; 17: 275–290.
27. Lin TH, Yao G. Evaluating Item Discrimination Power of WHOQOLBREF from an Item Response Model Perspectives. Soc Indic Res 2009; 91:141–153.
28. DeMars C. Item response theory. USA: Oxford University Press;2010.

Published by "Tehran University of Medical Sciences" (www.tums.ac.ir)