Psychol Stud
https://doi.org/10.1007/s12646-022-00645-y

ASSESSMENT

Psychometric Assessment of the 18-Item Bangla Mental Health Inventory (Bangla MHI-18)

Oli Ahmed[1] · **Mark D. Griffiths**[2] · **Muhammad Alamgir Hossain**[1]

Received: 18 January 2020 / Accepted: 24 January 2022

Abstract

The Mental Health Inventory-18 (MHI-18) is an 18-item self-report psychometric scale that assesses mental health among general populations. There are several adapted and translated mental health assessment scales in Bangladesh. However, none of these assess the complete state of mental health that the MHI assesses. Consequently, the present study assessed the psychometric properties of the MHI-18 in Bangladesh context (N=571). Confirmatory factor analysis confirmed the four-factor structure of the MHI-18 Bangla (i.e., anxiety, depression, behavior control, and positive affect). Second-order tests of measurement invariance suggested that the Bangla MHI-18 assessed the same construct in both male and female groups. The instrument had accepted levels of internal consistency reliability, test-retest reliability, composite reliability, average variance extracted, standard error of measurement, discrimination power, convergent validity, and concurrent validity. The tests relating to differential item functioning suggested lack of item response bias. Accepted levels of infit and outfit MnSqs, item and person separation, and reliabilities in rating scale model suggested that the Bangla MHI-18 had satisfactory psychometric properties. This will be helpful to Bangladeshi mental health practitioners to assess the complete mental state of young clients and contribute in formulating necessary therapeutic intervention for ensuring sound mental health.

Keywords: Mental health, confirmatory factor analysis, measurement invariance, reliability, validity, differential item function, Rasch analysis.

Psychometric assessment of the 18-item Bangla Mental Health Inventory (Bangla MHI-18)

The World Health Organization (WHO: 2014) defines health as "a state of complete physical, mental and social well-being, and not merely the absence of disease or infirmity" (p. 1) and is paramount to psychological wellbeing. The WHO also states that mental health is "a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community" (WHO, 2004; p. 10). It includes both psychological wellbeing and psychological distress. Studies suggest that components of mental health such as distress are associated with aggression (Jaser et al., 2005), poor academic achievement (Alva & Reyes, 1999; Cunningham, Hurley, Foney, & Hayes, 2002), and depression (Martin, Kazarian, & Breiter, 1995).

Low self-esteem has frequently been examined as a contributing factor in mental health problems including anxiety, depression, antisocial personality disorder, and suicidal ideation (Boden, Fergusson, & Horwood, 2008; Schroever, Ranchor, & Sanderman, 2003).Studies have suggested that life satisfaction is positively associated with better mental health (McKnight, Huebner, & Suldo, 2002), and that hope is negatively correlated with mental health problems and psychopathology (Hagen, Myers, & Machintosh, 2005; Snyder et al., 1997; Valle, Huebner, & Carolina, 2004). Headey, Kelley, and Wearing (1993) also suggested life satisfaction is an important dimension of mental health. However, this dimension has a stronger correlation with depression (Headey et al., 1993). Studies also suggest differences in mental health status in terms of socio-demographic factors (e.g., better mental status being male, employed, highly educated, and married) (Loge & Kaasa, 1998; Strand, Dalgard, Tambs, & Rognerud, 2003; Sullivan & Karlsson, 1998).

There are several mental health assessment tools that are currently used for assessing mental health status. Most of the assessment tools for mental health mainly focus on psychological distress, negative emotion, and symptom management (Keyes, 2007). Individuals prefer measures that help to express their concerns regarding the psychological wellbeing (Stedman, Yellowlees, Drake, Chant, Clarke, & Chapple, 2000). Clinical research has also focused on the positive aspects of mental health (Andresen, Caputi, & Oades, 2010; Westerhof & Keyes, 2010).

Mental health assessment tools should include both positive (psychological wellbeing) and negative (psychological distress) components that reflect the complete state of mental health

(Mandersheid, Ryff, Freeman, McKnight, Dhingra, & Strine, 2010). One such instrument is the Mental Health Inventory (MHI: Veit & Ware, 1983). This inventory was developed in response to early mental health tools that were very heterogeneous in content (Macmillan, 1957; Langner, 1962) and only included measures of health problems (both physical and psychosomatic symptoms), and symptoms of anxiety and depression. The MHI was specifically developed for the assessment of mental health in general populations. The inventory comprises two components – psychological wellbeing and psychological distress, and 38 items which encompass a wide range of both positive and negative emotion.

The MHI was originally developed for the Rand Health Insurance Corporation. Veit and Ware (1983) analyzed the data from the Rand Health Insurance Experiment and developed the measure. Factor structure of MHI was determined by exploratory and confirmatory factor analysis, and results suggested a hierarchical factor model. Lower order structure comprised five factors (anxiety, depression, emotional ties, general positive affect, and loss of behavioral and emotional control) and higher order structure comprised two factors (psychological distress and wellbeing) (Veit & Ware, 1983). However, later MHI studies on reported a different factorial structure. Some studies supported the two-factor structure (Ostroff, Woolverton, Berry, & Lesko, 1996; Heubeck & Neill, 2000) and some supported a five-factor structure (Tanaka & Huba, 1984; Manne& Schnoll, 2001). Ostroff et al (1996) and Heubeck and Neill (2000) assessed the psychometric properties of MHI on adolescents. Both studies suggested girls reported more psychological distress and less psychological wellbeing than boys. Hennessy, Patrick, and Swinbourne (2018) reassessed psychometric properties of the original MHI on an adult Australian sample and reported females had higher levels of psychological distress than males.

The 18-item Mental Health Inventory (MHI-18) is the shorter form of this scale and also has a hierarchical factor model. The lower order comprises four factors (anxiety, depression, behavioral control, and positive affect) and higher order comprises two factors (psychological distress and wellbeing). Meybodi, Saeedi, Behjati, Noorbala, Dastbaravardec, and Enjedany (2011) assessed the psychometric properties of the Farsi version of the MHI-18. They reported the measure as a highly reliable and structured instrument for assessing mental health of general populations. Their study supported the two-factor structure of the MHI-18 (i.e., psychological distress and wellbeing).

There are several translated and adapted Bangla measures for assessing mental health – the General Health Questionnaire (GHQ: Goldberg, 1978; Bangla version of GHQ-28: Banoo, 2001; Bangla version of GHQ-12: Ilyas & Ayesha, 2001), the Strengths and Difficulties Questionnaire (SDQ: Goodman, 1997; Bangla version: Mullick & Goodman, 2001), and the Warwick-Edinburgh Mental Well-being Scale (WEMWBS: Tennant et al., 2007; Bangla version: Rahman & Imran, 2013). However, none of these assess the complete state of mental health that the MHI assesses. Consequently, the main objective of the present study was to assess the psychometric properties of the Bangla MHI-18.

Method

Participants

The study sample comprised university students from the University of Chittagong (Bangladesh) selected utilizing a convenience sampling technique (N=571; 48.7% males and 51.3% females). Participants' mean age was 20.84 years (SD=1.97 years). Among participants, 34.9% resided in university accommodation and 65.1% resided outside of the university. Four-fifths were raised in a nuclear family (79%) and the remainders were raised in an extended family (21%).

Measures

All respondents completed a questionnaire booklet that included the Bangla MHI-18 (translation procedure described below in the 'Procedure' section) and the General Health Questionnaire-12 (GHQ-12: Goldberg, 1978; Bangla version: Ilyas & Ayesha, 2001), and a personal information form that included age, gender, place of residence (living in university accommodation or not), and family type.

The Mental Health Inventory-18: The 18-item version of the MHI is included in the Multiple Sclerosis Quality of Life Inventory (MSQLI: Ritvo, Fischer, Miller, Andrews, Paty, & LaRocca, 1997) and is a reasonably brief, reliable measure for assessing mental health among general populations. Each item is answered on a six-point Likert type scale ranging from 1 (*none of the time*) to 6 (*all of the time*). Scores are computed using the formula: score = [(mean total/subscale score – 1)x100]/5. If respondents skip nine or more items, then the total score should not be computed. If respondents skip two or more items in any subscale, than the subscale score should not be computed. Total/subscale scores range from 0 to 100 and higher scores indicate better

mental health. For subscales, higher scores indicate lower anxiety and depression, and higher behavioral control and positive affect. In the present study, the Cronbach alphas ranged from .81 to .94, stability coefficients ranged from .56 to .64, and inter-factor correlations ranged from -.70 to .93.

The General Health Questionnaire (GHQ-12): The GHQ-12 was used to assess the convergent validity of the Bangla MHI-18. The GHQ-12 comprises 12 items (six positive and six negative items). Each item on the GHQ-12 is answered on a four-point Likert type scale ranging from 1 (*not at all*) to 4 (*much more than usual*). In the present study, the Cronbach alpha of the Bangla GHQ-12 was .76 (95% CI [.69, .81]).

Procedure

Following the rules and guidelines suggested by the International Test Commission (ITC, 2018) for translation and adaptation of measurement instruments, the MHI-18was translated. At first, available literature on the Mental Health Inventory (MHI) published in different scientific journals were reviewed to ensure the MHI-18 had the same meaning in Bangladesh culture. Experts' opinions were sought to determine the equivalence of constructs between two cultures. Based on expert opinion and a literature review, it was concluded that the measure had same meaning and definitions and equally was applicable to Bangladesh culture.

Following this, the MHI-18 was translated into Bangla separately by two bilingual experts. They selected the best words and expressions while translating the measure. Their translations were then synthesized into one version. This synthesized translation was examined by two subject experts. They were requested to check the conceptual equivalence of words or phrases, but not a word-for-word translation. They recommended a few modifications of words and expressions. Then, the translated final draft was back-translated by two other bilingual experts and synthesized into one. Again, two subject experts compared the content of the original version and back-translated version of the MHI-18. They then rated the items of the two versions to ensure they had the same content. This translated measure was then administered to a sample of 40 students for piloting. This ensured that the translation was easily understandable to participants.

Data were collected utilizing the aforementioned questionnaire booklet. Each booklet took approximately 10-12 minutes to complete. The research team distributed the questionnaire booklet to 700 university students in classroom settings and all participants were given clear instructions

about what they had to do. A total of 571 participants returned their completed booklet (response rate = 81.6%). One month was taken to collect all the data.

Statistical analysis

The IBM Statistical Package for Social Science (SPSS) version 25.0, IBM AMOS 24.0 version, Microsoft Excel 2010, DIFAS (Differential Item Functioning Analysis System) 5.0 (Penfiled, 2013), jMetrik were used to analyze the data. Descriptive statistics (e.g., frequency, percentages, mean, standard deviation, skewness, and kurtosis) were used to assess the characteristics of participants. Normality of the data was assessed through skewness and kurtosis values. A skewness value larger than 2 and a kurtosis value larger than 7 suggests the non-normality of the data for large sample (N>300) (Kim, 2013). Psychometric properties of the MHI-18 Bangla were assessed using both classical test theory (CTT) and item response theory (IRT).

Before the main analyses, corrected item-total correlations were examined. An item-total correlation value less than .3 indicates that the item does not correlate very well with the scale overall (Field, 2017). Under CTT, the construct validity was assessed through a series of factor analyses. Exploratory factor analysis is more appropriate in the development phase of a psychological test or measure and confirmatory factor analysis is more appropriate where a model already established (Floyd & Widaman, 1995). The present study utilized confirmatory factor analysis over exploratory factor analysis because the structure models of the MHI-18 were already available in the literature. In confirmatory factor analysis, eight hypothetical models were tested. These models were – (i) one-factor model, (ii) two-factor correlated model, (iii) two-factor second order model, (iv) bifactor model of two factors, (v) four-factor correlated model, (vi) four-factor second order model, (vii) bifactor model with four factors, and (viii) hierarchical model (four factors at lower end and two factors at higher end).

Model fits were assessed through the χ^2/df ratio, comparative fit index (CFI), goodness-of-fit index (GFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). The χ^2/df ratio less than 2 suggests a good model fit (Bollen, 1989). However, a ratio value less than 5 is acceptable (Marsh & Hocevar, 1988). CFI, GFI, and TLI values greater than .95 are considered as evidence of good model fit (Hu & Bentler, 1999). For acceptable fit, these values must be above .90 (Dimitrov, 2012). RMSEA and SRMR values of .05 or less suggest good model fit (Brown & Cudeck, 1993; Byrne, 2016). RMSEA can be misleading when the sample size is not large and degrees of freedom are small.

Marsh, Hau, and Wen (2004) suggested less stringent criteria for an acceptable model fit (RMSEA≤.08, CFI≥.90, and TLI≥.90). Factor loadings from CFA were used to assess the average variance extracted (AVE) and composite reliability (CR). The recommended cutoff for the AVE is >.5 and CR is >.6 (Bagozzi & Yi, 1988).

Multigroup CFA was carried out for assessing the measurement invariance of the MHI-18 Bangla across gender. Measurement invariance provides information about whether a measure has same psychometric properties across independent groups. Measurement invariance was tested utilizing seven models: configural invariance (model 1), first-order factor loading invariance (model 2),first-order and second-order factor loading invariance (model 3), first-order, second-order factor loading and intercepts of measured variables invariance (model 4), first-order, second-order factor loading, and intercepts of assessed variables and first-order factor invariance (model 5), first- and second-order factor loadings, intercepts, and disturbances of first order-factor invariance (Model 6), and first- and second-order factor loadings, intercepts, disturbances of first-order factors, and residual variances of measured variables invariance (Model 7). In the Jöreskog tradition, non-significant $\Delta\chi^2$ suggested measurement invariance across groups (Byrne, 2010). However, this fit statistic is an excessive stringent test of invariance (MacCallum, Roznowski, & Necowitz, 1992). Cheung and Rensvold (2002) suggested using ΔCFI to make decisions concerning invariance. Chen (2007) recommended ΔCFI >-.010, ΔRMSEA<.015, and ΔSRMR<.01 as evidence of measurement invariance.

For reliability assessment, Cronbach alpha (α .≥ 7 expected; Nunnally, 1978), split-half reliability through Spearman-Brown formula (values between .70-.80 are viewed as sufficient; Furr, 2011), test-retest reliability (r =.7 minimum satisfying; Kline, 2015), standard error of measurement (a value less than SD/2 is expected; Wuang, Su, & Huang, 2012), and discriminatory power (Ferguson δ ≥.90 expected; Kline, 2015) were assessed. Convergent validity was estimated through correlations among the Bangla MHI-18 and its subscales. Concurrent validity was assessed through correlation between GHQ-12 and Bangla MHI-18.

Item response bias assessed using a DIF test. Mantel-Haenszel χ^2, Standardized Liu-Agresti Cummulative Common Log-Odds Ratio (LOR Z), and Standardized Cox's Noncentrality Parameter (COX Z) were calculated. Mantel-Haenszel χ^2 value ≥3.84 (p<.05), and LOR Z and COX Z values greater than from -2 to +2 suggest the presence of DIF (Penfield, 2013). Moreover, positive values of Liu-Agresti Cummulative Common Log-Odds Ratio and Cox's Noncentrality

Parameter estimator suggest DIF in favor of reference group and negative values of these statistics favor the focal group.

The rating scale model (RSM) was used to assess the item validity, item and person separation reliability and index in the Rasch analysis. Item validity was tested using weighted fit statistics (infit) mean square (MnSq) and outlier sensitive fit statistic (outfit) mean square. Infit MnSq and outfit MnSq values close to 1 indicates good model-data fit. Values less than 1 indicate responses are too consistent with what are expected from the model and results in a better quality of model than it actually is (Bond & Fox, 2007). Values greater than 1 indicate more variation in the data than expected (Bond & Fox, 2007). Accepted range of infit MnSq and outfit MnSq values is .5 to 1.5 (Linacre, 2012). In RSM, item and person reliability values larger than .7 are desirable (Chang et al., 2014). Item and person separation indices should be larger than 2 (Meyer, 2014).

Additionally, participants' distribution in terms of their mental health status assessed by the scale was computed. The cut-off scores suggested by Yuvaraj, Poornima, and Rashmi (2016) were used. Mental health scores of 61 or lower were considered as poor mental health, scores between 62 and 66 were considered as better mental health with some problems, and scores of 67 or above considered as good mental health.

Ethics

In the present study, all procedures were performed following the 1964 Helsinki Declaration and its later amendments and the study was approved by the research team's university ethics committee. The importance and purposes of this study were explained to respondents. They were assured about the confidentiality of their responses. Respondents had the right to withdraw their data at any stage of the study.

Results

Table 1 shows the mean, standard deviation, minimum, maximum, skewness, and kurtosis of each item of the Bangla MHI-18. Skewness and kurtosis values suggested normality of the data. Table 2 shows that corrected item-total correlations ranged from .365 (Item 3) to .686 (Item 15). All of the items had satisfactory level of corrected *item-total correlation*. Table 3 shows that Model 1 (one-factor model) had the worst model fit. Model fits of Models 2, 3, and 4 were also not at the accepted level. Model fits of final four models were at the accepted level. Among these, Model 7

(bifactor model with four factors) had the best model fit (χ^2/df= 2.561, *CFI* = .967, *GFI* =.948, *TLI* =.956, *RMSEA* = .052, and *SRMR* = .032) compared to any other models. However, items' regression weights (ranging from .219 to .720 for general factor and from .161 to .799 for primary factors) were lower than Models 5, 6, and 8. After examining the model fits and item regression weights, Model 6 was retained. This model had comparatively the best model fits (χ^2/df= 2.666, *CFI* = .960, *GFI* =.939, *TLI* =.953, *RMSEA* = .054, and *SRMR* = .039) among remaining three models and also had satisfactory item regression weights (ranging from .580 to .788). Table 2 also shows the item regression weights of males (ranging from .613 to .880) and females (ranging from .543 to .889).

Table 4 shows information regarding measurement invariance. In the configural model (Model 1), all parameters were freely estimated. This model had acceptable model fits (χ^2/df= 1.997, *CFI* = .953, *TLI* =.944, *RMSEA* = .042, and *SRMR* = .047). This table shows significant χ^2 changes at first-order factor loading invariance compared to configural model ($\Delta\chi^2$=27.347, p<.05), and first- and second-order factor loadings, intercepts, disturbances of first-order factors, and residual variances of measured variables invariance compared to first- and second-order factor loadings, intercepts, and disturbances of first-order factors invariance ($\Delta\chi^2$=40.915, p<.05).However, negligible changes in CFI, RMSEA and SRMR suggested invariance between male and female groups at all level of the model. Therefore, the Bangla MHI-18 assesses the same construct in both male and females.

Convergent and concurrent validity

Table 5 shows the convergent and concurrent validity of the Bangla MHI-18. It significantly correlated with anxiety (r=.790, p<.01, 95% CI [.757, .819]), depression (r=.487, p<.01, 95% CI [.422, .548]), behavior control (r=.666, p<.01, 95% CI [.618, .710]), and positive affect (r= .876, p<.01, 95% CI [.849, .889]) subscales. Table 5 also shows that GHQ-12 significantly correlated with Bangla MHI-18 (r=-.544, p<.01, 95% CI [-.653, -.413]) and its subscales (Anxiety = -.466, p<.01, 95% CI [-.589, -.323], depression = -394, p<.01, 95% CI [-.528, -.241], behavior control = -.492, p<.01, 95% CI [-.610, -.353], and positive affect = -.422, p<.01, 95% CI [-.552, -.272]).

Internal consistency reliability

Cronbach alphas of the MHI-18 and its subscales were good (MHI-18: .886, 95% CI [.872, .899]; anxiety: .854, 95% CI [.834, .872]; depression: .850, 95% CI [.829, .870]; behavior control:

.830, 95% CI [.805, .851]; and positive affect: .878, 95% CI [.861, .894]). Split-half reliabilities using the Spearman-Brown formula of this scale and its subscale were also satisfactory (MHI-18: .905, anxiety: .860, depression: .835, behavior control: .854, and positive affect: .868).

Test-retest reliability

Test-retest reliabilities were assessed on a sample of 45 university students over a one-month gap. Test-retest reliabilities of the MHI-18 Bangla version and its subscales were satisfactory (MHI-18: .925, 95% CI [.866, .958]; anxiety: .926, 95% CI [.869, .959]; depression: .904, 95% CI [.831, .946]; behavior control: .898, 95% CI [.821, .943]; and positive affect: .866, 95% CI [.861, .924]).

Average variance extracted (AVE)

The AVEs of the subscales of the MHI-18 Bangla version were all satisfactory (>.5) (anxiety: .54, depression: .59, behavior control: .65, and positive affect: .55).

Composite reliability (CR)

The composite reliabilities of the subscales of the MHI-18 Bangla version were .86for anxiety, .85 for depression, .88for behavior control, and .83for positive effect.

Standard error of measurement (SEM)

The *SEM* of the Bangla MHI-18 was 9.549. The subscales' SEM values were 7.823 (anxiety), 7.998 (depression), 9.168 (behavior control), and 7.556 (positive affect). All SEM values were at satisfactory level (<*SD*/2).

Discriminatory power

The Ferguson delta of the MHI-18 Bangla (.992) and its subscales (anxiety: .980, depression: .974, behavior control: .971, and positive affect: .964) were all at satisfactory level (>.90).

Differential item functioning (DIF)

Table 6 shows the DIF contrasts statistics of the MHI-18 across gender. Male was the reference group and female was the focal group. Mentel χ^2 statistics (ranging from .022 to 2.684), *Standardized Liu-Agresti Cumulative Common Log-Odds Ratio*s statistics (ranging from -1.653 to 1.519), and *Standardized Cox's Noncentrality Parameter*s (ranging from -1.636 to -1.565) suggested the non-DIF contrast between male and female. These findings suggested absence of item response bias in MHI-18 Bangla version between males and females.

Rasch Analysis

Table 6 also shows fit statistics of the Rasch analysis. All items infit MnSqs ranged from .80 (Item 9) to 1.26 (Items10 and 14) and outfit MnSqs are ranged from .78 (Item 9) to 1.25 (Item 10). Item separation indices ranged from 3.47 (behavior control subscale) to 11.86 (anxiety subscale), and person separation indices ranged from 1.95 (behavior control subscale) to 2.63 (depression subscale). Item reliabilities ranged from .92 (behavior control subscale) to .99 (anxiety and depression subscale), and person reliabilities from .79 (behavior control subscale) to .87 (depression subscale and positive affect subscale). The distribution of mental health status using the scale is presented in Table 7. Among the sample, 48% of participants' mental health status was rated as poor, and 39.4% of participants' mental health status was rated as good. The remainder of the participants (12.6%) had better mental health but with some problems.

Discussion

The Mental Health Inventory (MHI) is helpful in assessing an individual's state of mental health. The measure is useful in identifying intervention needs to ensure individuals are fully functional. The MHI not only focuses on alleviation of psychological distress but also on increased psychological wellbeing. The present study was undertaken to assess the psychometric properties of the Bangla MHI-18 for facilitating the mental health assessment of young Bangladeshi adults. Table 2 demonstrated that all items had sufficient item-total correlations. Discrimination index statistics provided information on whether a specific item and total test or part of the test battery assessed the same psychological construct. Item-total correlation provides the same sort of information. A positive item-total correlation indicates that the item is able to discriminate sufficiently between low scorers and high scorer in the test. Because all items of the Bangla MHI-18 had satisfactory item-total correlations, items of this measure are able to discriminate between high scorers (better mental health) and low scorers (poorer mental health).

The present study finding contributes to the existing contradictions concerning the aforementioned studies examining the factor structure of the full version of the MHI. The MHI-18 contains items assessing two negative psychological constructs (anxiety and depression which assess psychological distress) and two positive psychological constructs (behavior control and positive affect which assess psychological wellbeing). The present study confirmed that these four factors also existed in a Bangladeshi context. It also supported studies that explored five factors in original MHI (Tanaka & Huba, 1984; Manne & Schnoll, 2001). Meyabodi et al (2011) explored

two factors in the Farsi version of the MHI-18 via exploratory factor analysis. They reported the MHI-18 as a positively correlated bi-dimensional (psychological wellbeing and psychological distress) measure for Iranian university students. These two factors accounted for 63% of the variance. Studies conducted on adolescent samples suggest a two-factor model (Ostroff et al.,1996; Heubeck & Neill, 2000) and studies on relatively older adults suggest a five-factor model (Tanaka & Huba, 1984; Manne & Schnoll, 2001). A recent study conducted on Australian adult sample suggested a three-factor structure of the MHI in its original form (Hennessy et al., 2018). However, they excluded 17 items because these had insufficient factor loading or were cross-loaded.

Correlation coefficients in Table 5 demonstrated that correlations among MHI-18 and its four subscales were significantly correlated with each other (ranging from .210 to .874). Table 4 also showed that the MHI-18 and its four subscales were significantly and negatively correlated with the GHQ-12. A higher score on the MHI-18 and a lower score on the GHQ-12 both indicate better mental health. Consequently, the negative correlation between these two measures was expected. These data suggest that the MHI-18 has convergent validity. This result concerning convergent validity supports the utility of the Bangla MHI-18 to assess mental health and both its positive and negative aspects. Meybodi et al (2011) also found negative significant correlation between the Farsi version of MHI-18 and GHQ-28. Manne and Schnoll (2001) reported that positive affect subscale of the Positive and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988) had significant inverse correlations with anxiety, depression, and loss of behavioral/emotional control, and a strong positive correlation with general positive affect, and significant positive correlation with emotional ties.

Cronbach alphas, split-half reliabilities, test-retest reliabilities, and composite reliabilities of the MHI-18 Bangla were all at satisfactory level. Nunnally (1978) reported that a reliability of .70 or higher would be expected before using the instrument. Kline (1999) suggested that an alpha less than .5 is unacceptable, .5 to .6 is poor, .6 to .7 is acceptable, .7 to .9 is good, and above .9 is excellent. Based on these recommendations, the reliabilities of the Bangla MHI are good. Meybodi et al (2011) reported a Cronbach alpha of .93 for the Farsi version of MHI-18, and the subscales Cronbach alphas ranging from .63 to .85. The split-half reliability, in their study was, .93. Heuback and Neill (2000) reported high Cronbach alphas (.94 for psychological distress and .92 for psychological wellbeing) and sufficient test-retest reliability (.71 for psychological distress, .69 for psychological wellbeing, and .73 for total MHI). Average variance extracted (AVE) and

standard error of measurement (SEM) values were also above the minimum criteria. Significant correlation with GHQ-12, internal consistency reliabilities, test-retest reliabilities, composite reliabilities, AVEs, and SEMs of the MHI-18 Bangla all suggest that the scale has sufficient reliability and validity.

Information regarding DIF contrast between males and females suggested the absence of item response bias across gender. None of the items of this measure favored either male or female. Fit statistics of Rasch analysis (Table 6) suggest that all items had sufficient item fit statistics (infit MnSq and outfit MnSq). None of the items' fit statistics exceeded the acceptable range suggested by Linacre (2012). However, person separation index of the behavior control subscale was less than (but close to) the accepted level (>2). Separation indices suggested that the scale has the ability to separate items of individuals into two or more groups (Kook & Varni, 2008). However, item and person reliabilities of the Bangla MHI-18 were above the minimum acceptable level (>.7). Item reliability suggests the reproducibility in the order of item difficulty and person reliability suggest the reproducibility of persons' underlying ability (Chang et al., 2014).

Using the scale, the present study also profiled the mental health status of all participants. Results suggested that nearly half of the participants' mental health condition was poor while approximately two-fifths of the participants' mental health condition was good. In a countrywide survey, the National Institute of Mental Health (NIMH; 2019) reported the prevalence of any mental disorder to be 16.8%. This prevalence rate is much lower than the present study results. However, the present study only included only young adults. Kessler et al. (2005) reported higher prevalence of any mental disorder among US young adults (52.4%). Gustavson et al. (2018) found that the prevalence of any mental disorder among Norwegian young adults was 27.8%. The present study assessed the distribution of mental health status based on the cut-off of another study that was conducted in India. Therefore, this is another possible reason for this discrepancy between study findings. Moreover, a self-report scale never gives the complete picture of an individual's mental health but merely provides some indications of possible mental health problem symptoms. Based on the findings of the present study, it is concluded that nearly half of the participants had poor mental health symptoms but that this is indicative rather than being definitive.

Utility of the Bangla Mental Health Inventory-18

The psychometric results of the scale under investigation suggested the inventory is a reliable and valid instrument. The Bangla MHI-18 has both research and practical utility. The inventory can assess both positive aspects (behavioral control and positive affect) and negative aspects (anxiety and depression) of mental health among Bangladeshi young adults. Therefore, the inventory is able to provide a broad indicative assessment of respondents' mental health status. The Bangla MHI-18 meets the need of the Bangladeshi researchers who want to assess both aspects of mental health at the same time. This inventory lessens the burden of using separate psychological scales or tests to assess these aspects of mental health. Moreover, use of the instrument reduces survey fatigue among participants that occurs when researchers utilize longer scales or tests. In practice, the Bangla MHI-18 inventory will be helpful to mental health practitioners assessing mental health status among emerging adults in Bangladesh. Moreover, the inventory will help contribute in the formulating of necessary interventions for the alleviation of psychological distress and the strengthening of psychological wellbeing.

Limitations and future recommendations

The present study has some limitations which should be taken into account when interpreting the findings. A major limitation existed regarding the sample selection. The study sample only comprised university students aged from 19-26 years old who were selected using anon-representative sampling technique from single university of Bangladesh. Consequently, there was no clinical sample included. Another major limitation of this study was the unavailability of a screening norm of the MHI-18. Screening norms for this measure have not been established for Bangladeshi people. Moreover, the present study utilized the self-report data. Therefore, there might be possibilities of some well-known biases (such as social desirability bias, method bias, memory recall bias, etc.) regarding such data. However, some of these limitations could be overcome by conducting a study with a large representative sample that includes respondents from all ages and all groups. Such a study would be helpful in establishing the screening norms to identify individuals for whom professional therapeutic assistance would be needed. Despite these limitations, the findings demonstrate that the 18-item Bangla Mental Health Inventory is a valid and reliable instrument for assessing mental health among Bangladeshi young adults. The inventory has applicability in both research and practice and will provide indicative and reliable assessment of the mental health status of young Bangladeshi individuals.

Abbreviations

CFI: Comparative Fit Index

CTT: Classical Test Theory

COX Z: Standardized Cox's Noncentrality Parameter

DIF: Differential item functioning

GFI: Goodness-of-Fit Index

GHQ: General Health Questionnaire

IRT: Item Response Theory

ITC: International Test Commission

LOR Z: Standardized Liu-Agresti Cummulative Common Log-Odds Ratio

MHI: Mental Health Inventory

MnSq: Mean Square

MSQLI: Multiple Sclerosis Quality of Life Inventory

RMSEA: Root Mean Square Error of Approximation

RSM: Rating Scale Model

SDQ: Strengths and Difficulties Questionnaire

SRMR: Standardized Root Mean Square Residual

TLI: Tucker–Lewis Index

WEMWBS: Warwick-Edinburgh Mental Well-being Scale

WHO: World Health Organization

References

Alva, S. A., & Reyes, R. d. L. (1999). Psychosocial stress, internalized symptoms, and the academic achievement of Hispanic adolescents. *Journal of Adolescent Research, 14*(3), 343-358. doi: 10.1177/0743558499143004

Andresen, R., Caputi, P., & Oades, L. (2010). Do clinical outcome measures assess consumer-defined recovery? *Psychiatry Research, 177*(3), 309–317. doi: 10.1016/j.psychres.2010.02.013

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 16*(1), 74–94. doi:10.1007/BF02723327

Banoo, N. (2001). *The translated Bangla version of GHQ-28*. Department of Clinical Psychology, University of Dhaka.

Boden, J. M., Fergusson, D. M., & Horwood, L. J. (2008). Does adolescent self-esteem predict later life outcomes? A test of the causal role of self-esteem. *Development and Psychopathology*, *20*(1), 319–339. doi: 10.1017/S0954579408000151

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd edition). New York: Routledge.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd edition). New York: Routledge.

Chang, K.-C., Wang, J.-D., Tang, H.-P., Cheng, C.-M., & Lin, C.-Y. (2014). Psychometric evaluation, using Rasch analysis, of the WHOQOL-BREF in heroin-dependent people undergoing methadone maintenance treatment: Further item validation. *Health and Quality of Life Outcomes, 12*(1), 148. doi:10.1186/s12955-014-0148-6

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504. doi:10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. doi:10.1207/S15328007SEM0902_5

Cunningham, M., Hurley, M., Foney, D., & Hayes, D. (2002). Influence of perceived contextual stress on self-esteem and academic outcomes in African American adolescents. *Journal of Black Psychology*, *28*(3), 215–233. doi: 10.1177/0095798402028003003

Dimitrov, D. M. (2012). *Statistical methods for validation of assessment scale data in counseling and related fields*. Alexandria, VA, US: American Counseling Association.

Field, A. (2017). *Discovering statistics using SPSS* (5th edition). London: Sage Publications.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286-299. doi: 10.1037/1040-3590.7.3.286

Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. Thousand Oaks, CA: Sage Publication Ltd.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*(5), 581-586. doi: 10.1111/j.1469-7610.1997.tb01545.x

Goldberg, D. (1978). *Manual of the General Health Questionnaire.* Windsor: NFER-Nelson.

Gustavson, K., Knudsen, A. K., Nesvåg, R., Knudsen, G. P., Vollset, S. E., & Reichborn-Kjennerud, T. (2018). Prevalence and stability of mental disorders among young adults: Findings from a longitudinal study. *BMC Psychiatry*, *18*(1), 65. https://doi.org/10.1186/s12888-018-1647-5

Hagen, K. A., Myers, B. J., & Machintosh, V. H. (2005). Hope, social support, and behavioral problems in at-risk children. *American Journal of Orthopsychiatry*, *75*(2), 211–219. doi: 10.1037/0002-9432.75.2.211

Headey, B. W., Kelley, J., & Wearing, A. J. (1993). Dimensions of mental health: Life satisfaction, positive affect, anxiety and depression. *Social Indicators Research, 29*(1), 63-82. doi: 10.1007/BF01136197

Hennessy, M. J., Patrick, J. C., & Swinbourne, A. L. (2018). Improving mental health outcomes assessment with the Mental Health Inventory-21. *Australian Psychologist*, *53*(4), 313-324. doi:10.1111/ap.12330

Heubeck, B. G., & Neill, J. T. (2000). Confirmatory factor analysis and reliability of the Mental Health Inventory for Australian adolescents. *Psychological Reports*, *87*(2), 431–440. doi: 10.2466/pr0.2000.87.2.431

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

Ilyas, Q. S. M., & Aeysha, M. (2001). *Bangali version of the Goldberg General Health Questionnaire (GHQ). Unpublished Manuscript*. Department of Psychology: University of Dhaka.

International Test Committee (2018). The ITC guidelines for translating and adapting tests (second edition). *International Journal of Testing*, *18*(2), 101-134. doi: 10.1080/15305058.2017.1398166

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593-602. https://doi.org/10.1001/archpsyc.62.6.593

Keyes, C. L. M. (2007). Promoting and protecting mental health as flourishing. *American Psychologist, 62*(2), 95–108. doi: 1037/0003-066X.62.2.95

Kline, P. (1999). *The handbook of psychological testing* (2nd edition). London: Routledge.

Kline, P. (2015). *A handbook of test Construction: Introduction to psychometric design.* New York: Routledge.

Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics, 38*(1), 52-54. doi: 10.5395/rde.2013.38.1.52

Kook, S. H., & Varni, J. W. (2008). Validation of the Korean version of the pediatric quality of life inventory™ 4.0 (PedsQL™) generic core scales in school children and adolescents using the Rasch model. *Health and Quality of Life Outcomes, 6*(1), 41. doi:10.1186/1477-7525-6-41

Jaser, S. S., Langrock, A. M., Keller, G., Merchant, M. J., Benson, M. A., Reeslund, K., … Compas, B. E. (2005). Coping with the stress of parental depression II: Adolescent and parent reports of coping and adjustment. *Journal of Clinical Child and Adolescent Psychology*, *34*(1), 193–205. doi: 10.1207/s15374424jccp3401_18

Langner, T. S. (1962). A twenty-two item screening score of psychiatric symptoms indicating impairment. *Journal of Health & Human Behavior, 3*(4), 269-276. doi: 10.2307/2948599

Linacre, J. M. (2012). *A user's guide to WINSTEPS and ministep Rasch model computer programs: Program manual. 3.75*. Retrieved from https://www.winsteps.com/manuals.htm

Loge, J. H., & Kaasa, S. (1998). Short form 36 (SF-36) health survey: normative data from the general Norwegian population. *Scandinavian Journal of Social Medicine*, *26*(4), 250–258. doi: 10.1177/14034948980260040401

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance

structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490-504. doi: 10.1037/0033-2909.111.3.490

Macmillan, A. M. (1957). The Health Opinion Survey: Technique for estimating prevalence of psychoneurotic and related types of disorder in communities. *Psychological Reports, 3*(2), 325-339. doi: 10.2466/PR0.3..325-339

Mandersheid, R. W., Ryff, C. D., Freeman, E. J., McKnight, L. R., Dhingra, S., & Strine, T. W. (2010). Evolving definitions of mental illness and wellness. *Prevention of Chronic Disease, 7*(1), 1–6.

Manne, S., & Schnoll, R. (2001). Measuring cancer patients' psychological distress and well-being: A factor analytic assessment of the Mental Health Inventory. *Psychological Assessment, 13*(1), 99-109. doi: 10.1037/1040-3590.13.1.99

Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology, 73*(1), 107-117. doi: 10.1037/0021-9010.73.1.107

Marsh, H.W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fi t indexes and dangers in over-generalizing findings. *Structural Equation Modeling, 11*(3), 320–341. doi: 10.1207/s15328007sem1103_2

Martin, R. A., Kazarian, S. S., & Breiter, H. J. (1995). Perceived stress, life events, dysfunctional attitudes, and depression in adolescent psychiatric inpatients. *Journal of Psychopathology and Behavioral Assessment*, *17*(1), 81–95. doi: 10.1007/BF02229205

McKnight, C. G., Huebner, E. S., & Suldo, S. (2002). Relationships among stressful life events, temperament, problem behavior, and global life satisfaction in adolescents. *Psychology in the Schools*, *39*(6), 677–687. doi: 10.1002/pits.10062

Meybodi, F. A., Saeedi, Z., Behjati, Z., Noorbala, F., Dastbaravardec, A., & Enjedany, E. (2011). Reliability and validity of a Farsi version of 18-item mental health inventory. *Procedia - Social and Behavioral Sciences*, *30*, 1425–1429. doi: 10.1016/j.sbspro.2011.10.277

Mullick, M. S. I., & Goodman, R. (2001). Questionnaire screening for mental health problem in Bangladeshi children: A preliminary analysis. *Social Psychiatry and Psychiatric Epidemiology, 36*(2), 94-99. doi: 10.1007/s001270050295

National Institute of Mental Health (2019). *National Mental Health Survey of Bangladesh, 2018-19: Provisional fact sheet.* National Institute of Mental Health.

Nunnally, J. C. (1978). *Psychometric Theory* (2nd edition). New York: McGraw-Hill.

Ostroff, J. S., Woolverton, K. S., Berry, C., & Lesko, L. M. (1996). Use of the mental health inventory with adolescents: A secondary analysis of the Rand Health Insurance study. *Psychological Assessment, 8*(1), 105–107. doi: 10.1037/1040-3590.8.1.105

Penfiled, R. D. (2013). *DIFAS 5.0. Differential Item Functioning Analysis System. Users's Manual*. Retrieved May 21, 2019, from: https://soe.uncg.edu/wp-content/uploads/2015/12/DIFASManual_V5.pdf

Rahman, S. T., & Imran, M. A. (2013). Bangladeshi adaptation of Warwick-Edinburgh Mental Well-being scale. *The Dhaka University Journal of Psychology, 37*, 49-60.

Ritvo, P. G., Fischer, J. S., Miller, D. M., Andrews, H., Paty, D. W., & LaRocca, N. G. (1997). *Multiple Sclerosis Quality of Life Inventory: A user's manual*. NewYork: National Multiple Sclerosis Society.

Schroever, M. J., Ranchor, A. V., & Sanderman, R. (2003). Depressive symptoms in cancer patients compared with people from the general population: The role of sociodemographic and medical factors. *Journal of Psychosocial Oncology*, *21*(1), 1–26. doi: 10.1300/J077v21n01_01

Snyder, C. R., Hoza, B., Pelham, W. E., Rapoff, M., Ware, L., Danovsky, M., … Stahl, K. J. (1997). The development and validation of the Children's Hope Scale. *Journal of Pediatric Psychology*, *22*(3), 399–421. doi: 10.1093/jpepsy/22.3.399

Stedman, T., Yellowlees, P., Drake, S., Chant, D., Clarke, R., & Chapple, B. (2000). The perceived utility of six selected measures of consumer outcomes proposed for routine use in Australian mental health services. *Australian and New Zealand Journal of Psychiatry, 34*, 842–849. doi: 10.1080/j.1440-1614.2000.00803.x

Strand, B. H., Dalgard, O. S., Tambs, K., & Rognerud, M. (2003). Measuring the mental health status of the Norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nordic Journal of Psychiatry*, *57*(2), 113-118. doi: 10.1080/08039480310000932

Sullivan, M., & Karlsson, J. (1998). The Swedish SF-36 Health Survey III. Evaluation of criterion-based validity: Results from normative population. *Journal of Clinical Epidemiology*, *51*(11), 1105–1113. doi: 10.1016/S0895-4356(98)00102-4

Tanaka, J. S., & Huba, G. J. (1984). Structures of psychological distress: Testing confirmatory

hierarchical models. *Journal of Consulting and Clinical Psychology, 52*(4), 719-721. doi: 10.1037/0022-006X.52.4.719

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., …..& Stewart-Brown, S. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health & Quality of Life Outcomes*, *5*(63), 114212. doi:10.1186/1477-7525-5-63

Valle, M. F., Huebner, E. S., & Suldo, S. M. (2004). Further evaluation of the Children's Hope Scale.*Journal of Psychoeducational Assessment, 22*(4), 320–337. doi: 10.1177/073428290402200403

Veit, C. T., & Ware, J. E. (1983). The structure of psychological distress and well-being in general populations. *Journal of Consulting and Clinical Psychology*, *51*(5), 730–742. doi:10.1037/0022-006X.51.5.730

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070. doi: 10.1037/0022-3514.54.6.1063

Westerhof, G. J., & Keyes, C. L. M. (2010). Mental illness and mental health: The two continua model across the lifespan. *Journal of Adult Development*, *17*(2), 110–119. doi: 10.1007/s10804-009-9082-y

World Health Organization. (2004). *Promoting mental health: Concepts, emerging evidence, practice: Summary report*. Geneva: World Health Organization.

World Health Organization (2014). *Basic documents* (48th edition). Geneva: World Health Organization.

Wuang, Y. P., Su, C. Y., & Huang, M. H. (2012). Psychometric comparisons of three measures for assessing motor functions in preschoolers with intellectual disabilities. *Journal of Intellectual Disability Research, 56*(6), 567–578. doi:10.1111/j.1365-2788.2011.01491.x

Yuvaraj, Y., Poornima, S., & Rashmi, S. (2016). Screening for overall mental health status using mental health inventory amongst medical students of a government medical college in North Karnataka, India. *International Journal of Community Medicine and Public Health*, *3*(12), 3308-3312. https://doi.org/10.18203/2394-6040.ijcmph20164026

Tables

Table 1. The descriptive statistics of the Bangla MHI-18

Item	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Item 1	4.12	1.30	1.00	6.00	-.67	-.35
Item 2	3.50	1.28	1.00	6.00	.27	-.70
Item 3	4.82	1.27	1.00	6.00	-1.14	.68
Item 4	3.63	1.28	1.00	6.00	.09	-.68
Item 5	4.53	1.33	1.00	6.00	-.82	-.10
Item 6	3.46	1.31	1.00	6.00	.18	-.64
Item 7	4.26	1.18	1.00	6.00	-.63	-.11
Item 8	4.32	1.29	1.00	6.00	-.74	-.01
Item 9	3.81	1.25	1.00	6.00	.12	-.90
Item 10	3.75	1.31	1.00	6.00	-.39	-.54
Item 11	3.64	1.24	1.00	6.00	.09	-.73
Item 12	3.57	1.23	1.00	6.00	.06	-.70
Item 13	4.40	1.18	1.00	6.00	-.74	-.09
Item 14	4.24	1.21	1.00	6.00	-.45	-.61
Item 15	4.37	1.38	1.00	6.00	-.76	-.31
Item 16	4.44	1.42	1.00	6.00	-.52	-.83
Item 17	4.65	1.43	1.00	6.00	-.80	-.39
Item 18	3.53	1.30	1.00	6.00	.16	-.62

Table 2. Corrected item-total correlations, factor weights of the MHI-18 Bangla

Items	Corrected Item total correlations	Factor weights of CFA	Factor weights across gender	
			Male	Female
Anxiety: *m* = 52.06, *SD* = 20.47				
Item 4	.645	.696	.655	.737
Item 6	.706	.766	.779	.746
Item 10	.582	.654	.668	.646
Item 11	.683	.754	.747	.764
Item 18	.724	.810	.837	.778
Depression: *m* = 55.59, *SD* = 20.65				
Item 2	.721	.812	.809	.818
Item 9	.751	.837	.842	.832
Item 12	.721	.790	.814	.760
Item 14	.580	.629	.613	.641
Behavior control: *m* = 69.68, *SD* = 22.23				
Item 5	.650	.723	.704	.751
Item 8	.610	.780	.837	.740
Item 16	.654	.836	.779	.889
Item 17	.715	.875	.880	.860
Positive Affect: *m* = 67.92, *SD* = 20.13				
Item 1	.685	.711	.732	.707
Item 7	.724	.649	.773	.543
Item 13	.764	.751	.720	.762
Item 15	.788	.844	.793	.875
Mental Health: *m* = 61.17, *SD* = 15.07				

Table 3. Model fit statistics of the Bangla MHI-18 in confirmatory factor analysis

Fit statistics	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
χ^2/df	16.220	15.241	15.241	11.371	2.705	2.666	2.561	3.411
CFI	.583	.650	.650	.780	.960	.960	.967	.942
GFI	.643	.681	.681	.780	.939	.939	.948	.924
TLI	.527	.597	.597	.706	.952	.953	.956	.932
RMSEA	.163	.158	.158	.135	.055	.054	.052	.065
sRMR	.129	.131	.131	.109	.038	.039	.032	.083

Model 1 - unidimensional model, Model 2 - two dimensional correlated model, Model 3 - two dimensional second order model, Model 4 - two dimensional bifactor model, Model 5 - four dimensional correlated model, Model 6 - four dimensional second order model, Model 7 - four dimensional bifactor model, and Model 8 – hierarchical model (four factors at lower end and two factors at higher end)

Table 4. Summary of fit statistics for testing measurement invariance of MHI-18 Bangla

Models	χ^2	CFI	RMSEA	sRMR	Model Comparison	$\Delta\chi^2$	ΔCFI	ΔRMSEA	ΔsRMR
Model 1	457.932	.953	.042	.047					
Model 2	485.279	.950	.042	.051	M2-M1	27.347*	.003	0	.004
Model 3	488.072	.950	.042	.053	M3-M2	2.793	0	0	.002
Model 4	507.761	.950	.040	.053	M4-M3	19.689	0	.002	0
Model 5	508.868	.950	.040	.053	M5-M4	1.107	0	0	0
Model 6	513.359	.949	.040	.054	M6-M5	4.491	.001	0	.001
Model 7	554.274	.944	.041	.057	M7-M6	40.915*	.005	-.001	.003

*p<.05

CFI: comparative fit index; RMSEA: root mean square error of approximation; sRMR: standardized root mean square residual.

Model 1 = configural model, Model 2 = first order factor loading invariant, Model 3 = first order and second order factor loading invariant, Model 4 = first order, second order factor loading and intercepts of measured variables invariant, Model 5 = first order, second order factor loading, and intercepts of measured variables and first order factors invariant, Model 6 = first- and second-order factor loadings, intercepts, and disturbances of first-order factors invariant, Model 7 = first- and second-order factor loadings, intercepts, disturbances of first-order factors, and residual variances of measured variables invariant

Table 5. Correlation coefficients among the Bangla MHI-18 and its four subscales, and GHQ-12

Variables	Mental Health	Anxiety	Depression	Behavior Control	Positive Affect
Anxiety	.790**				
	(.757, .819)				
Depression	.487**	.234**			
	(.422, .548)	(.155, .310)			
Behavior Control	.666**	.449**	.210**		
	(.618, .710)	(.381, .512)	(.130, .287)		
Positive Affect	.876**	.596**	.230**	.410**	
	(.855,.894)	(.540, .646)	(.151 .306)	(.340, .476)	
GHQ-12 (n=135)	-.544**	-.466**	-.394**	-.492**	-.422**
	(-.653, -.413)	(-.589, -.323)	(-.528, -.241)	(-.610, -.353)	(-.552, -.272)

**p<.01, Values in the parenthesis are 95% confidence interval

Table 6. Differential item functioning statistics and Rasch analysis fit statistics of Bangla MHI-18

Items	DIF statistics			Rasch analysis fit statistics					
	Mantel χ^2	LOR Z	COX Z	IfM	OfM	ISI	PSI	IR	PR
Anxiety									
Item 4	.056	.238	.237	1.05	1.05	11.86	2.18	.99	.83
Item 6	.355	-.597	-.594	.91	.91				
Item 10	1.204	1.11	1.094	1.26	1.25				
Item 11	.018	-.134	-.131	.90	.89				
Item 18	.746	-.863	-.867	.85	.84				
Depression									
Item 2	.627	.799	.789	.97	.95	9.98	2.63	.99	.87
Item 9	.239	-.487	-.488	.80	.78				
Item 12	.339	-.597	-.584	.91	.86				
Item 14	.036	.189	.191	1.26	1.25				
Behavior control									
Item 5	1.854	-1.368	-1.358	.99	.96	3.47	1.95	.92	.79
Item 8	1.858	1.318	1.36	1.04	1.01				
Item 16	.148	-.381	-.388	1.04	.97				
Item 17	.145	.393	.378	.96	.86				
Positive Affect									
Item 1	.357	-.597	-.595	1.19	1.16	3.90	2.58	.94	.87
Item 7	.003	-.006	-.059	.97	.95				
Item 13	.250	-.493	-.498	.85	.86				
Item 15	1.547	1.268	1.242	.99	.96				

Reference group = Male, Focal group = Female;

L-A LOR = *Liu-Agresti cummulative common log-odds ratio*, LOR SE = *Standard error of the Lui-Agresti Cumulative Common Log-Odds Ratio*, LOR Z = *Standardized Liu-Agresti Cummulative Common Log-Odds Ratio*, COX's B = *Cox's Noncentrality Parameter Estimator*, COX SE = *Standard Error of Cox's Noncentrality Parameter Estimator*, COX Z = *Standardized Cox's Noncentrality Parameter*

IfM = infit MnSq, OfM = outfit MnSq, ISI = Item separation index, PSI = Person separation index, IR = item reliability, PR = person reliability

Table 7. Distribution of mental health status

	Overall mental health
Poor mental health	48.0%
Better mental health with some problems	12.6%
Good mental health	39.4%