Psikohumaniora: Jurnal Penelitian Psikologi
Vol 8, No 2(2023): 295–308. DOI: 10.21580/pjpp.v8i2.18077
https://journal.walisongo.ac.id/index.php/Psikohumaniora
ISSN 2527-7456 (online); 2502-9363 (printed)

Assessing adaptation of the psychometric properties of the General Health Questionnaire (GHQ-12) using the Rasch Measurement Model among Indian teenagers

Syed Faraz Ali,[1] Aqeel Khan,[1]* Adibah Binti Abdul Latif,[1] Abdul Wahab Pathath,[2] Shabnam Shabnam,[3] Mohammad Asif,[4] Arieff Salleh Rosman[5]

[1]School of Education, Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia, Johor – Malaysia; [2]Department of Clinical Neurosciences, College of Medicine, King Faisal University, Al-Ahsa – Saudi Arabia; [3]Department of Humanities and Social Sciences, National Institute of Technology, Kurukshetra, Haryana – India; [4]Department of Psychology, Faculty of Social Sciences, Aligarh Muslim University, Aligarh – India; [5]Center of Research for Fiqh Science and Technology (CFiRST), Universiti Teknologi Malaysia, Johor – Malaysia

Abstract: The General Health Questionnaire (GHQ-12) is widely accepted as an instrument used to examine diagnosable psychiatric disorders. It is also routinely used as a measure of psychological morbidity. This study aimed to assess and establish the psychometric properties of the GHQ-12 among Indian teenagers. The tool was administered to 212 adolescents aged from 12 to 18 years. Following two-stage cluster sampling, the data were collected from Aligarh district, India. Fleiss Kappa analysis was used to determine test reliability and showed an overall value of 0.94, based on the rater agreement for the instrument. Furthermore, the Rasch measurement model was used, with values of 0.79 and 0.83 for person and item reliability, respectively. Moreover, the value of unidimensionality was found to be 37.9%. Additionally, item fit statistics and item analysis were conducted for the instrument. Based on the preliminary data and findings, the study provides primary evidence for the reliability and validity of GHQ-12. Hence, the questionnaire can lead to more multi-site studies in India.

Keywords: General Health Questionnaire; instrument validation; measuring tool validation

Abstrak: *General Health Questionnaire* (GHQ-12) adalah alat ukur standar yang umum digunakan untuk menilai kecenderungan permasalahan psikologis. Instrumen ini juga dapat digunakan untuk mengukur morbiditas atau kerentanan psikologis seseorang. Penelitian ini bertujuan untuk mengetahui atribut psikometrik dari alat ukur GHQ-12 yang diberikan kepada 212 remaja berusia 12-18 tahun di Aligarh, India. Analisis Fleiss Kappa menunjukkan nilai reliabilitas tes sebesar 0,94. Berikutnya, dari analisis instrumen menggunakan *Rasch Model* didapatkan hasil sebesar 0,79 untuk *person reliability* dan 0,83 untuk *item reliability*. Selain itu, didapatkan indeks unidimensionalitas sebesar 37,9%. Selanjutnya, statistik *goodness-of-fit* dan analisis item juga dilakukan untuk instrumen tersebut. Penelitian ini menghasilkan pembuktian terkait reliabilitas dan validitas instrumen GHQ-12. Oleh karena itu, GHQ-12 dapat digunakan pada lebih banyak penelitian di berbagai wilayah di India.

Kata Kunci: *General Health Questionnaire*; validasi instrument; validasi alat ukur

*__Corresponding Author:__ Aqeel Khan (aqeel@utm.my), Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia, 81310 Skudai, Johor – Malaysia.

Submitted: 27 Sep 2023; Received in revised form: 25 Nov 2023; Accepted: 26 Nov 2023; Published regularly: Nov 2023

Introduction

Good health and well-being are contingent upon a solid mental health foundation (Khan et al., 2023) and thus influence the social and economic outcomes of our entire lives (Herrman & Jané-Llopis, 2012; Jenkins et al., 2011). Children and adolescents especially need a strong basis for healthy development and good mental health (Barry et al., 2013), and now, adolescents' account for almost 16% of the world'spopulation —1.3 billion individuals, as those between the ages of 10 and 19— (UNICEF, 2023). At present, the collective load of mental health-related problems among the adolescent and teenager population is rising and has become a concern across the world (Bor et al., 2014). Numerous mental health disorders commence before the age of 25, and more typically during the adolescent years or in young teenagers (Patton et al., 2016). The load connected with common mental disorders (such as depression and anxiety disorders) rises in childhood and was found to be highest in adolescence and early teenage groups (Chadda, 2018). A meta-analysis study projected a global incidence of general mental disorder among children and adolescents of 13.4% (Polanczyk et al., 2015). Hence, it is important to assess the mental health of this targeted population to help reduce their chances of the condition developing into a disorder.

The General Health Questionnaire (GHQ) is widely accepted and used as an instrument to examine diagnosable psychiatric disorders (Goldberg & Hillier, 1979). The original version of the GHQ comprised 60 items (GHQ-60) and was subsequently modified into reduced versions, namely the GHQ-30, GHQ-28 (Creed, 2023), GHQ-20, and GHQ-12 (Goldberg & William, 1988). The 12-item GHQ-12 is the most widely used version and helps to provide an insight into the general and most common mental disorders, along with an individual's current mental health status. The GHQ-12 has achieved global recognition and has been translated into multiple languages, including English, French, German, and Hindi, thus demonstrating significant validity in both clinical and general population contexts (Werneke et al., 2000). A cumulative GHQ score generated by aggregating the responses from all items represents an individual's overall level of psychological distress.

The GHQ-12, as one of the most widely adopted instruments for gauging common psychological distress conditions, was developed by Goldberg (1972) and comprises 12 items that gauge the severity of mental health problems among individuals. The questionnaire is essentially structured around four sub-constructs, namely anxiety (items 1–3), depression (items 4–6), social dysfunction (items 7–9), and loss of confidence (items 10–12). Each item is evaluated on a four-point rating scale encompassing 'less than usual,' 'no more than usual,' 'rather more than usual,' and 'much more than usual.' It produces a total score in the range of 0–36, based on the scoring method employed by the researcher. While the GHQ-12 can be scored in various ways, for example, standard GHQ scoring, modified Likert scoring, or chronic scoring, the most prevalent scoring methods are Bi-model (0-0-1-1) and Likert rating scoring (0-1-2-3). It is recommended that the psychometric evaluation of the GHQ-12 is tested prior to use, and it should have high reliability when measured (C. R. Martin & Jomeen, 2003). Moreover, numerous studies from different countries have reported good psychometric analysis results for the GHQ-12, particularly in terms of its reliability and validity. For instance, the global score of internal consistency reliability for the GHQ-12 typically ranges from 0.79 to 0.91 (Hankins, 2008b; Shevlin & Adamson, 2005); however, a composite reliability score of 0.90 was also recorded across a variety of scoring methods (Rey et al., 2014). Likewise, test-retest reliability yielded a value of

0.84 after 7 to 14 days (Piccinelli et al., 1993), dipping slightly to 0.79 after 20 days (López-Castedo & Fernández, 2005), and was expected to decline even further as the interval period lengthened (Quek et al., 2001). Moreover, the value of validities across 17 studies displayed a moderate sensitivity of 0.84 and a specificity of 0.79 (Goldberg et al., 1997; C. R. Martin & Newell, 2005).

While the GHQ-12 is widely employed and is considered a well-validated and highly reliable instrument (Werneke et al., 2000), the literature contains various ongoing debates, notably concerning the dimensionality analysis of the instrument (Smith et al., 2010). Focusing on the structure of the GHQ-12, although the original instrument focuses solely on a unidimensionality test, a considerable volume of literature based on the evidence of exploratory and confirmatory factor analysis suggests that the GHQ-12 has a two- or three-factor structure (A. J. Martin, 1999; Schrnitz et al., 1999). Additionally, aside from the two- or three-factor framework suggested, the significant notable correlations observed between the factors have frequently prompted various researchers to advocate for the use of aggregated GHQ-12 scores (French & Tait, 2004; Gao et al., 2004; Shevlin & Adamson, 2005). However, when examining the individual GHQ-12 items, they are categorized into positively and negatively phrased items and if employed on a large sample population, the data are steadier, consistent, and more reliable when assessed using a one-dimensional measurement approach (Hankins, 2008a). However, since the original version of the GHQ-12 was conceived as a unidimensional measure, considerable controversy has surrounded the factor structure that underlies it. Consequently, only a limited number of studies have confirmed the single-factor structure (Fernandes & Vasconcelos-Raposo, 2013). It is therefore crucial to discuss and thoroughly evaluate the inherent unidimensionality of the original GHQ-12 as indicated, in addition to its reliability and validity. Based on the reliability and validity result, the questionnaire is expected to meet the requirement of measuring what it intends to measure with much greater consistency. Hence, it is also important to measure the other aspects to ensure that the instrument meets the requirements of the study and is thus widely accepted.

Therefore, this study focused on assessing the reliability and validity of the GHQ-12 in the context of an Indian sample. While previous studies have provided evidence of its consistency and reliability when employed on the general population of India (Kashyap & Singh, 2017; Mangal et al., 2020; Mohanan et al., 2012), this study aimed to further investigate its applicability to the Indian population in response to the growing demand for its use. As such, the primary objective of this study was to evaluate the test reliability, that is, to analyze and report inter raters' agreement using Fleiss Kappa analysis, as well as to discuss the item and person reliability and unidimensionality using the Rasch measurement model for the GHQ-12 among the Indian teenager population. The Rasch measurement model was chosen as it typically provides a framework that enables researchers to compare empirical data to evaluate how well the instrument replicates the fundamental measurement properties (notably invariance and unidimensionality) and can therefore be used as a tool for quantifying human conditions that are not observable.

Method

Participants and Sampling

For this study, the survey instrument was administered to a population of 212 Indian teenagers ranging from 11 to 18 years of age. The respondent data were collected from various sources, including school settings spanning classes from 6th to 12th, coaching centers, and local communities. The current study employed a two-stage cluster sampling method specifically utilizing

"disproportional random sampling" within the district of Aligarh, India. The district was divided into four zones: North, South, East, and West, referred to as "clusters," to facilitate the collection of data. Then, based on the disproportional sampling method, in subsequent stages, respondents were selected from each zone based on the study requirement and participant availability. Moreover, the respondents were chosen randomly from different schools, coaching centers, and local communities. The data were collected from the participants using a physical questionnaire, completed with paper and pencil. The participants were assured that their personal details would not be shared or exposed to anyone at any cost and would thus remain confidential. Initially, the participants were asked to fill out their demographic details, including gender, age, class, family type (nuclear/joint), and school type (government/private).

Rasch Analysis

The Rasch model is a type of latent trait model used to assess person measures and item difficulties across a single continuum. It identifies and establishes the probability relationship between item difficulty and person ability, both of which are represented as "logits" or log-odds units (Rasch, 1960). Logits are commonly used to assess individuals or items among the latent continuum. In the event of dichotomous item responses, they are derived from the natural logarithm of the probability of scoring 1 over the probability of scoring 0. The resulting logits then represent the difference between the location of the person and the location of the items. There are two important categories for Rasch measurement: item fit and dimensionality. Item fit is typically evaluated using the mean square residual fit statistic (Bond & Fox, 2015). Within this category, the primary statistics for assessing item fit are the infit and outfit statistics. While these were anticipated to be around 1.0, acceptable values fall within the range of 0.5–1.5. The infit statistic detects residuals close to the estimated person ability, while the outfit statistic identifies outliers for either person or item parameters (Bond & Fox, 2015). Dimensionality, meanwhile, considers whether a single-factor model can effectively capture the entire variance in the data, essentially addressing whether the instrument is unidimensional or not. This can be estimated through principal component analysis (PCA) of the residuals once the Rasch factor has been extracted (Bond & Fox, 2015; Smith et al., 2010). Moreover, Rasch (Rasch, 1960) suggested that the validity and reliability of the items in a research instrument can also be determined using the Rasch measurement model.

Analysis can be performed using Winsteps software, which includes person reliability, item reliability, item fit, item measure, and person fit statistics. The reliability value can be seen in the person and item reliabilities in Tables 4 and 5, respectively. Moreover, Rasch modeling involves the estimation of item and person parameters, including item thresholds. It is important to acknowledge the potential difficulty in interpreting the meaning of these thresholds in practical terms. The Rasch model essentially comprises two popular measurement models: The Rating Scale Model (RSM), developed by Andrich (2019), and the Partial Credit Model (PCM) by Masters (1982). This study employed the RSM, whereby items have the same response options, and the thresholds, representing the deviance of the particular category from the overall item difficulty or severity, remain the same across items. Hence, items can have different overall item difficulty severities.

General Health Questionnaire (GHQ-12)

The GHQ, designed by Goldberg and Willaims in 1988, is used to assess current mental health well-being by evaluating normal health functioning and the presence of distressing symptoms. As a self-

report measure of psychological morbidity, it is intended to detect psychiatric disorders in community and non-psychiatric settings (Goldberg & William, 1988). The GHQ-12 has undergone extensive evaluation concerning the severity of psychological morbidity (Hahn et al., 2006; Navarro et al., 2007; Quek et al., 2001). The instrument employs a four-point response scale to indicate the severity of symptoms; 'not at all,' 'same as usual,' 'rather more than usual,' or 'much more than usual.' It can be scored in either a binary format (0-0-1-1) or a four-point Likert response format (0-1-2-3), giving a potential score range of 0–36.

Statistical Analysis

The study encompassed a comprehensive analysis of the participants' demographic characteristics and the prevalence rates of the GHQ-12 were summarized. The participants' demographic profile was compiled, in which they were asked to state their age, gender, family type (nuclear/joint), and school type (government/private). This information contributed to a holistic understanding of the sample composition. In terms of inferential statistics, Fleiss Kappa analysis was conducted based on the input of four expert reviewers (McHugh, 2012), who evaluated the questionnaire items based on the relevancy and clarity score. This facilitated the measurement of the questionnaire's test reliability, providing insights into the degree to which the items aligned with the intended construct, which was based on rater agreement with very limited information and not on empirical evidence. Regarding the construct validity, unidimensionality analysis was conducted using the Rasch measurement model. This analytical process served to evaluate the ability of the instrument to consistently measure a single latent trait across its terms. For reliability, Winsteps software was used to determine the person and item reliability. The two types of reliability were analyzed separately as they served different purposes in the Rasch modeling. Person reliability is more important in terms of understanding the reliability and stability of the scores for individuals, which is particularly important in applications such as education and clinical assessment. Item reliability, meanwhile, helps in evaluating and improving the quality of the assessment tool itself, ensuring that the items are well-constructed and provide consistent information. Furthermore, SPSS version 23 was used to facilitate the statistical analysis. This enhanced the accuracy and precision of the data analysis, enabling robust insights to be drawn from the collected information.

Result

Test Reliability

For the test reliability analysis, Fleiss Kappa analysis was performed based on the expert rating and reviews. The instrument was sent to four examiners who were experts in the field of Psychology and Education. They made their comments based on the relevance and clarity of the questionnaire items. The rating was based on a four-point Likert scale. Table 1 shows the Fleiss Kappa analysis.

Table 1 indicates that the sub-constructs Depression, Social Dysfunction, and Loss of Confidence had the highest Kappa values (K=0.95), thus demonstrating almost perfect agreement. However, every sub-construct had an almost perfect agreement. The findings were consistent across all sub-constructs, further underscoring the robustness of the questionnaire content. The GHQ had an overall kappa value of 0.94, while the value of all sub-constructs individually exceeded the substantial agreement value ($K \geq 0.40$) among the experts (Gisev et al., 2013; Latif, 2013); hence, the questionnaire was suitable for study. It can thus be concluded with confidence that all the items in the questionnaire are well accepted for inclusion in the study.

Table 1

Fleiss Kappa for the GHQ-12

Sr. No	Sub-Construct	Kappa Value	Interpretation
1.	Anxiety	0.91	Almost Perfect Agreement
2.	Depression	0.95	Almost Perfect Agreement
3.	Social Dysfunction	0.95	Almost Perfect Agreement
4.	Loss of Confidence	0.95	Almost Perfect Agreement
	Overall Fleiss Kappa	0.94	Almost Perfect Agreement

Table 2

Item Statistics for the GHQ-12

Item	Item Measure	Standard Error	Infit MNSQ	Outfit MNSQ	PTMEA Correlation
D8	-1.16	.30	1.37	1.38	.29
D3	-.34	.30	1.35	1.36	.44
D1	-.34	.30	1.17	1.18	.55
D2	1.24	.30	1.10	1.11	.52
D7	.37	.30	1.08	1.11	.38
D11	.72	.29	1.07	1.08	.57
D9	.55	.30	.98	.99	.69
D5	.98	.29	.91	.90	.65
D10	.11	.30	.86	.88	.66
D6	-.43	.30	.76	.75	.63
D4	-1.16	.30	.70	.69	.71
D12	-.52	.30	.63	.63	.62
Mean	.00	.30	1.00	1.00	
S.D.	.76	.00	.23	.23	

These results reinforce the effectiveness of the questionnaire in capturing the intended constructs and enhance its credibility as a robust research tool.

Item Fit: Misfit Order

Item fit, assessed through item statistics, is usually estimated by iterative evaluation involving the removal of misfitting items and those with negative polarity. The remaining items are then systematically adjusted and re-evaluated until no further instances of negative polarity or misfit items are identified. This process ensures that the instrument is refined to a high degree of precision, eliminating items that may introduce bias into the data collection. Consequently, the instrument's reliability and accuracy in assessing the intended constructs are enhanced, contributing to the robustness of the study findings. Fortunately, this study contained no negative polarity of item misfit to remove. It could thus be concluded that the questionnaire was highly precise and suitable for use with the population. Table 2 illustrates the item fit: misfit order for the GHQ-12. In this study, no item was omitted while performing the analysis.

Dimensionality

For construct validity, unidimensionality analysis was conducted using the Rasch measurement model. This model employs unidimensionality, which means it focuses on one trait or dimension at a time in assessing construct validity (Bond & Fox, 2015). A difference of 20% is the basis

for unidimensionality; that is, unidimensionality requires a difference in raw performance of at least 20%, as defined by the measures. Furthermore, the undefined difference of the first construct should not exceed 15% (Bond & Fox, 2015). Reckase (1979) similarly suggested that raw variance explained by measures of greater than 20% is also acceptable. Meanwhile, Sumintono and Widhiarso (Sumintono & Widhiarso, 2015) reported that the interpretation values for unidimensionality based on raw variance explained by measures were accepted if equal to or above 20%, good if equal to or above 40%, and excellent if equal to or above 60%, while the Eigenvalue for the first contrast could not exceed 5 (Bond & Fox, 2015; Linacre, 2005). The unidimensionality of the GHQ-12 for this study is presented in the Tables 3.

Table 3 indicates the unidimensionality of the GHQ-12. The value of unidimensionality is 37.9%; this exceeds 20%, which means the value is accepted. The total variability in the observation is 19.3, representing 100% of the empirical evidence, indicating that the Rasch model captures the entirety of the observed variance. The Eigenvalue of the first contrast, that is, the error, is 2.3, which is less than 5; it is therefore accepted. Moreover, the measure explains 19.5% of the variance in a person's response. This reflects the extent to which the Rasch model captures individual differences in the latent trait. In addition, the measure explains 18.5% of the variance in the individual items that is accounted for by the latent trait. These results collectively indicate the effectiveness of the Rasch measurement model in explaining a substantial portion of the variance in the GHQ-12 data, while acknowledging the presence of some unexplained variance, particularly in specific contrast. This nuanced understanding enhances the interpretability of the model's performance in assessing the unidimensionality of the instrument. Hence, the instrument sub-constructs are considered valid (Linacre, 2005; Reckase, 1979).

Table 4 shows a person reliability value of 0.79 for the GHQ-12. Since this value is above 0.6, person reliability is acceptable (Bond & Fox, 2015). The mean person measure is 32.4, indicating the average ability level of the participants, while the standard deviation is 4.3, reflecting the variability in the distribution of person measure. The infit and outfit mean squares (MNSQ) are indices of person-item fit. The infit and outfit MNSQ both have values of 1.00 and lie within an acceptable range, suggesting a good fit between the respondents and the items. Moreover, the questionnaire has a person separation value of 1.92; this is also considered satisfactory as it exceeds the minimum required value of 1.5 (Souza et al., 2017), thereby denoting the instrument's ability to differentiate between individual varying levels of the latent trait. Higher values indicate better discriminatory power. Hence, the GHQ-12 short version had acceptable person reliability and separation. Moreover, it should be mentioned that no item was omitted from the GHQ to obtain these acceptable values. Thus, the person fit statistic suggests that the final version of the GHQ-12 questionnaire demonstrated satisfactory psychometric properties. The mean and standard deviation of the person measure, along with the fit indices and reliability measure, collectively contribute to a comprehensive assessment of the instrument's performance in measuring the latent trait of the study and the overall psychological health among adolescents in Aligarh district, India.

Table 5 shows that the GHQ-12 had an item reliability value of 0.83, indicating the consistency and stability of the item measures. Values closer to 1.0 suggest a more reliable measure. Since the value is above 0.6, item reliability is acceptable (Bond & Fox, 2015). The mean person measure is 108.1, indicating the average ability level of the participants. The standard deviation (SD) is 8.5, reflecting the variability in the distribution of person measures. The infit and outfit MNSQ are indices of person-item fit. Infit MNSQ is 0.00, which

suggests a perfect fit, and outfit MNSQ is 1.00, which also suggests a perfect fit to the Rasch model. Additionally, the questionnaire had person and item separation values of 2.20, respectively; this is also considered satisfactory as it exceeds the minimum required item separation value of 1.5 (Linacre, 2005; Souza et al., 2017). Hence, the item fit statistics suggest that the GHQ-12 demonstrates excellent psychometric properties. The mean and SD of the measure, along with the fit indices and reliability measures, collectively contribute to a comprehensive assessment of the instrument's performance in measuring the latent trait. Moreover, it should be mentioned that no item was omitted from the GHQ to obtain these acceptable reliability values. Thus, the final version of the GHQ includes 12 items to study the overall psychological health among adolescents in Aligarh district, India.

Table 3

Unidimensionality of the General Health Questionnaire (GHQ-12)

		Empirical		Modeled
Total raw variance in observations	19.3	100.0%		100.0%
Raw variance explained by measures	7.3	37.9%		37.9%
Raw variance explained by persons	3.8	19.5%		19.5%
Raw variance explained by items	3.6	18.5%		18.4%
Raw unexplained variance (total)	12.0	62.1%	100.0%	62.1%
Unexplained variance in 1st contrast	2.3	11.9%	19.2%	
Unexplained variance in 2nd contrast	1.9	9.6%	15.5%	
Unexplained variance in 3rd contrast	1.4	7.1%	11.4%	
Unexplained variance in 4th contrast	1.3	7.0%	11.2%	
Unexplained variance in 5th contrast	1.1	5.8%	9.4%	

Table 4

Person Reliability of the General Health Questionnaire

Person	212 Inputs		198 Measured		INFIT		OUTFIT	
	Total	Count	Measure	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
Mean	32.4	12.0	.73	.59	1.00	.0	1.00	.0
SD	4.3	.0	1.28	.05	.38	1.1	.39	1.1
Real RMSE	.59	TRU SD	1.13	Separation	1.92	Person Reliability .79		

REALSE	Real Standard Error	REAL RMSE	Real Root Mean Square Error
IMNSQ	Mean Square Infit Statistic	SD	Standard Deviation
ZSTD	Infit Mean Square Fit Statistic	INFIT	Information-Weighted Fit Statistic
OMNSQ	Mean Square Outfit Statistic	OUTFIT	Outlier-Sensitive Fit Statistic

Table 5

Person Reliability of the General Health Questionnaire

Person	12 Inputs		12 Measured		INFIT		OUTFIT	
	Total	Count	Measure	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
Mean	108.1	40.0	.00	.31	1.00	.0	1.00	.0
SD	8.5	.0	.76	.02	.23	1.0	.23	1.0
Real RMSE	.59	TRU SD	1.13	Separation	2.20	Item Reliability .83		

REALSE	Real Standard Error	REAL RMSE	Real Root Mean Square Error
IMNSQ	Mean Square Infit Statistic	SD	Standard Deviation
ZSTD	Infit Mean Square Fit Statistic	INFIT	Information-Weighted Fit Statistic
OMNSQ	Mean Square Outfit Statistic	OUTFIT	Outlier-Sensitive Fit Statistic

Discussion

The study aimed to assess the validity and reliability of the GHQ-12 as a well-known and widely recognized tool used in evaluating minor psychological distress and mental health status, which has been translated into many different languages all over the world (Daradkeh et al., 2001; Doi & Minowa, 2003; Politi et al., 1994; Quek et al., 2001). In this study, the GHQ-12 was administered to a sample population of Indian teenagers. While prior instances of GHQ-12 administration have been reported in India (Kashyap & Singh, 2017; Mangal et al., 2020; Mohanan et al., 2012), the growing demand for and application of the questionnaire necessitated a comprehensive examination of its reliability and validity across diverse regions of India. The analysis result confirms that the GHQ-12 has good reliability and validity. Reliability was tested using Fleiss Kappa analysis and yielded a value of 0.94, which is greater than 0.7 and is considered to be an 'almost perfect agreement' for the value (Gisev et al., 2013; Latif, 2013). This robust value supports the questionnaire's effectiveness.

Moreover, the Rasch measurement model was employed to comprehensively explore both the item and person reliability, along with the dimensional analysis. The analytical approach was conducted after items exhibiting negative polarity and item misfit had been excluded, as determined through the item statistics. The item and person reliability values were subsequently found to be 0.83 and 0.79, respectively, surpassing the threshold of 0.7 for both (Bond & Fox, 2015). This proved that the instrument has good reliability. Furthermore, the analysis extended to examining the dimensionality of the instrument. It achieved a unidimensionality value of 37.9% and an Eigenvalue of 2.3. These outcomes align with established standards of goodness and acceptability for this study, consistent with the criteria established (Linacre, 2005; Reckase, 1979). This underscores the instrument's efficacy in capturing the intended unidimensional construct, thus solidifying its value for assessing the targeted psychological dimensions. Moreover, the alignment of the results with those of previous studies demonstrates that the GHQ-12 shows good and acceptable values in the context of its use among a population of Indian adolescents.

Hence, the GHQ-12's performance was evaluated based on both reliability and validity criteria. The reliability assessment indicated a high level of consistency, which suggests that the questionnaire items consistently measured the same underlying construct. It included the determination of both item and person reliability. This rigorous examination ensured that the instrument reliably measured both the individuals and the items within the scores. In terms of validity, the GHQ-12 displays a significant correlation as another established mental health assessment tool, indicating that it effectively captures the intended psychological aspects. The GHQ-12 also shows a strong correlation value between the dimensions themselves. The result of this study supports the nature of the GHQ-12, while each of its sub-scales also correlates (Goldberg, 1972). The score indicates its fitness for study and the result supports those of previous studies conducted on both the Indian population and the general populations of other countries. Therefore, the findings from this study build on and strengthen the results for this instrument and can be used to examine general psychological distress.

However, it is important to also acknowledge the limitations inherent in the current study. A notable limitation lies in the relatively low number of participants, which suggests the potential for enhanced robustness through a larger sample size in future research endeavors. Additionally, the

researcher aspires to extend this line of investigation by conducting similar studies within the same population. This broader exploration will not only deepen our understanding but also shed light on the broader applicability and usefulness of the GHQ-12 across a wider spectrum. This expansion could provide a more comprehensive evaluation of the instrument's validity and reliability on a larger scale, potentially yielding more conclusive results.

Conclusion

The conventional psychometric analysis conducted in this study leads to the conclusion that the GHQ-12 shows suitable reliability and validity. The promising result demonstrates the potential for future multi-site studies that can further validate and enhance the use of the questionnaire. The implementation of a validated tool to measure general psychological distress can aid in developing appropriate interventions and policies for youth with various mental health problems, leading to more effective support for this vulnerable population. The reliability and validity analysis of the GHQ-12 makes a significant contribution to the research on general mental health disorders and provides a foundation for future studies. Based on the values obtained from the Fleiss Kappa analysis, dimensionality, item statistics, and the item and person reliability of the GHQ-12, it can be concluded that the instrument is fit for study and can be used for the sample population in the Indian context.

Generally, Rasch modeling relies on several assumptions about the study in question, including its independence, unidimensionality, and monotonicity. Violation of these assumptions generally creates biased parameters and hence, misinterpreted results. Therefore, it is essential to acknowledge that deviations from these assumptions may affect the validity of the findings. Moreover, Rasch analysis typically requires a large sample to obtain stable and reliable data, based on item and person parameters. However, for this study, discussion would have been needed concerning the potential impact of a limited sample size on the precision and generalizability of the result. Furthermore, the current study was limited to exploring certain fit indices for the analysis, such as rater consistency, standard error measurement, and item DIF (Differential Item Functioning).[]

Acknowledgment

The researchers are grateful to all the participants who took part in this study and all the stakeholders for their assistance. We are also grateful to the universities for their permission to conduct the study.

Author Contribution Statement

Syed Faraz Ali: Conceptualization; Formal Analysis; Methodology; Writing Original Draft; Writing Review & Editing. **Aqeel Khan:** Conceptualization; Writing, Formal Analysis; Methodology; Review. **Adibah Binti Abdul Latif:** Conceptualization; Formal Analysis; Methodology; Review. **Abdul Wahab Pathath:** Formal Analysis; Review. **Shabnam:** Formal Analysis; Review**. Mohammad Asif:** Formal Analysis. **Arieff Salleh Rosman:** Formal Analysis; Review.

References

Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory: Measuring in the educational, social and health sciences* (Vol. 41). Springer Nature Singapore. https://doi.org/10.1007/978-981-13-7496-8

Barry, M. M., Clarke, A. M., Jenkins, R., & Patel, V. (2013). A systematic review of the effectiveness of mental health promotion interventions for young people in low and middle income countries. *BMC Public Health*, *13*(1), 835. https://doi.org/10.1186/1471-2458-13-835

Bond, T., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental measurement in the human sciences*. Routledge. https://doi.org/10.4324/9781315814698

Bor, W., Dean, A. J., Najman, J., & Hayatbakhsh, R. (2014). Are child and adolescent mental health problems increasing in the 21st century? A systematic review. *Australian & New Zealand Journal of Psychiatry*, *48*(7), 606–616. https://doi.org/10.1177/0004867414533834

Chadda, R. (2018). Youth & mental health: Challenges ahead. *Indian Journal of Medical Research*, *148*(4), 359–361. https://doi.org/10.4103/ijmr.IJMR_1585_18

Creed, F. (2023). Psychiatric disorders comorbid with general medical illnesses and functional somatic disorders: The Lifelines cohort study. *Plos One*, *18*(5), e0286410. https://doi.org/10.1371/journal.pone.0286410

Daradkeh, T. K., Ghubash, R., & El-Rufaie, O. E. F. (2001). Reliability, validity, and factor structure of the Arabic version of the 12-item General Health Questionnaire. *Psychological Reports*, *89*(1), 85–94. https://doi.org/10.2466/pr0.2001.89.1.85

Doi, Y., & Minowa, M. (2003). Factor structure of the 12-item General Health Questionnaire in the Japanese general adult population. *Psychiatry and Clinical Neurosciences*, *57*(4), 379–383. https://doi.org/10.1046/j.1440-1819.2003.01135.x

Fernandes, H. M., & Vasconcelos-Raposo, J. (2013). Factorial validity and invariance of the GHQ-12 among clinical and nonclinical samples. *Assessment*, *20*(2), 219–229. https://doi.org/10.1177/1073191112465768

French, D. J., & Tait, R. J. (2004). Measurement invariance in the General Health Questionnaire-12 in young Australian adolescents. *European Child & Adolescent Psychiatry*, *13*(1), 1–7. https://doi.org/10.1007/s00787-004-0345-7

Gao, F., Luo, N., Thumboo, J., Fones, C., Li, S.-C., & Cheung, Y.-B. (2004). Does the 12-item General Health Questionnaire contain multiple factors and do we need them? *Health and Quality of Life Outcomes*, *2*(2), 63. https://doi.org/10.1186/1477-7525-2-63

Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, *9*(3), 330–338. https://doi.org/10.1016/j.sapharm.2012.04.004

Goldberg, D. P. (1972). *The detection of psychiatric illness by questionnaire: A technique for the identification and assessment of non-psychotic psychiatric illness*. Oxford University Press.

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine*, *27*(1), 191–197. https://doi.org/10.1017/S0033291796004242

Goldberg, D. P., & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine*, *9*(1), 139–145. https://doi.org/10.1017/S0033291700021644

Goldberg, D. P., & William, P. (1988). *A user's guide to the General Health Questionnaire*. Nfer-Nelson.

Hahn, D., Reuter, K., & Härter, M. (2006). Screening for affective and anxiety disorders in medical patients - comparison of HADS, GHQ-12 and Brief-PHQ. *Psycho-Social Medicine*, *3*, Doc09. http://www.ncbi.nlm.nih.gov/pubmed/19742274

Hankins, M. (2008a). The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing? *Clinical Practice and Epidemiology in Mental Health*, *4*(1), 10. https://doi.org/10.1186/1745-0179-4-10

Hankins, M. (2008b). The reliability of the twelve-item general health questionnaire (GHQ-12) under realistic assumptions. *BMC Public Health*, *8*(1), 355. https://doi.org/10.1186/1471-2458-8-355

Herrman, H., & Jané-Llopis, E. (2012). The status of mental health Promotion. *Public Health Reviews*, *34*(2), 6. https://doi.org/10.1007/BF03391674

Jenkins, R., Baingana, F., Ahmad, R., McDaid, D., & Atun, R. (2011). Social, economic, human rights and political challenges to global mental health. *Mental Health in Family Medicine*, *8*(2), 87–96. http://www.ncbi.nlm.nih.gov/pubmed/22654971

Kashyap, G. C., & Singh, S. K. (2017). Reliability and validity of general health questionnaire (GHQ-12) for male tannery workers: A study carried out in Kanpur, India. *BMC Psychiatry*, *17*(1), 102. https://doi.org/10.1186/s12888-017-1253-y

Khan, A., Sriyanto, S., Baranovich, D. L., Tahir, L. M., Panatik, S. A., Sasmoko, S., Indrianti, Y., Hamid, M. A. A., Lukman, F., Van, N. T., & Jabor, M. K. (2023). The relationship between positive mental health, academic stress and suicide ideation among Malaysian adolescents residing in Johor Bahru. *Current Psychology*, *42*(18), 15718–15726. https://doi.org/10.1007/s12144-022-02885-7

Latif, A. A. (2013). *Pembangunan instrumen penilaian akhlak pelajar di institusi pengajian tinggi awam*. [Doctoral dissertation]. Universiti Teknologi Malaysia.

Linacre, J. M. (2005). Standard errors: Means, measures, origins and anchor values. *Rasch Measurement Transactions: Archives of the Rasch Measurement SIG, AERA*, *19*(3), 1030. https://www.rasch.org/rmt/rmt193e.htm

López-Castedo, A., & Fernández, L. (2005). Psychometric properties of the Spanish version of the 12-Item General Health Questionnaire in adolescents. *Perceptual and Motor Skills*, *100*(3), 676–680. https://doi.org/10.2466/pms.100.3.676-680

Mangal, A., Thakur, A., Nimavat, K., Dabar, D., & Yadav, S. (2020). Screening for common mental health problems and their determinants among school-going adolescent girls in Gujarat, India. *Journal of Family Medicine and Primary Care*, *9*(1), 264. https://doi.org/10.4103/jfmpc.jfmpc_732_19

Martin, A. J. (1999). Assessing the multidimensionality of the 12-item General Health Questionnaire. *Psychological Reports*, *84*(3), 927–935. https://doi.org/10.2466/pr0.1999.84.3.927

Martin, C. R., & Jomeen, J. (2003). Is the 12-item General Health Questionnaire (GHQ-12) confounded by scoring method during pregnancy and following birth? *Journal of Reproductive and Infant Psychology*, *21*(4), 267–278. https://doi.org/10.1080/02646830310001622088

Martin, C. R., & Newell, R. J. (2005). Is the 12-item General Health Questionnaire (GHQ-12) confounded by scoring method in individuals with facial disfigurement? *Psychology & Health*, *20*(5), 651–659. https://doi.org/10.1080/14768320500060061

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/BF02296272

McHugh, M. L. (2012). Interrater reliability: The Kappa statistic. *Biochemia Medica*, *22*(3), 276–282. http://www.ncbi.nlm.nih.gov/pubmed/23092060

Mohanan, P., Jain, A., Shashidhar Kotian, M., & Vinay, N. K. (2012). Are the Anganwadi workers healthy and happy? A cross sectional study using the General Health Questionnaire (GHQ 12) at Mangalore, India. *Journal of Clinical & Diagnostic Research*, *6*(7), 1151–1154. https://doi.org/10.7860/JCDR/2012/.2464

Navarro, P., Ascaso, C., Garcia-Esteve, L., Aguado, J., Torres, A., & Martín-Santos, R. (2007). Postnatal psychiatric morbidity: A validation study of the GHQ-12 and the EPDS as screening tools. *General Hospital Psychiatry*, *29*(1), 1–7. https://doi.org/10.1016/j.genhosppsych.2006.10.004

Patton, G. C., Sawyer, S. M., Santelli, J. S., Ross, D. A., Afifi, R., Allen, N. B., Arora, M., Azzopardi, P., Baldwin, W., Bonell, C., Kakuma, R., Kennedy, E., Mahon, J., McGovern, T., Mokdad, A. H., Patel, V., Petroni, S., Reavley, N., Taiwo, K., … Viner, R. M. (2016). Our future: A Lancet commission on adolescent health and wellbeing. *The Lancet*, *387*(10036), 2423–2478. https://doi.org/10.1016/S0140-6736(16)00579-1

Piccinelli, M., Bisoffi, G., Bon, M. G., Cunico, L., & Tansella, M. (1993). Validity and test-retest reliability of the italian version of the 12-item General Health Questionnaire in general practice: A comparison between three scoring methods. *Comprehensive Psychiatry*, *34*(3), 198–205. https://doi.org/10.1016/0010-440X(93)90048-9

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*, *56*(3), 345–365. https://doi.org/10.1111/jcpp.12381

Politi, P. L., Piccinelli, M., & Wilkinson, G. (1994). Reliability, validity and factor structure of the 12-item General Health Questionnaire among young males in Italy. *Acta Psychiatrica Scandinavica*, *90*(6), 432–437. https://doi.org/10.1111/j.1600-0447.1994.tb01620.x

Quek, K. F., Low, W. Y., Razack, A. H., & Loh, C. S. (2001). Reliability and validity of the General Health Questionnaire (GHQ-12) among urological patients: A Malaysian study. *Psychiatry and Clinical Neurosciences*, *55*(5), 509–513. https://doi.org/10.1046/j.1440-1819.2001.00897.x

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. In *Studies in mathematical psychology*. Danmarks Paedagogiske Institut.

Reckase, M. D. (1979). Unifactor latent Trait Models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, *4*(3), 207–230. https://doi.org/10.3102/10769986004003207

Rey, J. J., Abad, F. J., Barrada, J. R., Garrido, L. E., & Ponsoda, V. (2014). The impact of ambiguous response categories on the factor structure of the GHQ–12. *Psychological Assessment*, *26*(3), 1021–1030. https://doi.org/10.1037/a0036468

Schrnitz, N., Kruse, J., & Tress, W. (1999). Psychometric properties of the General Health Questionnaire (GHQ-12) in a German primary care sample. *Acta Psychiatrica Scandinavica*, *100*(6), 462–468. https://doi.org/10.1111/j.1600-0447.1999.tb10898.x

Shevlin, M., & Adamson, G. (2005). Alternative factor models and factorial invariance of the GHQ-12: A large sample analysis using Confirmatory Factor Analysis. *Psychological Assessment*, *17*(2), 231–236. https://doi.org/10.1037/1040-3590.17.2.231

Smith, A. B., Fallowfield, L. J., Stark, D. P., Velikova, G., & Jenkins, V. (2010). A Rasch and confirmatory factor analysis of the General Health Questionnaire (GHQ) - 12. *Health and Quality of Life Outcomes*, *8*(1), 45. https://doi.org/10.1186/1477-7525-8-45

Souza, M. A. P., Coster, W. J., Mancini, M. C., Dutra, F. C. M. S., Kramer, J., & Sampaio, R. F. (2017). Rasch analysis of the Participation scale (P-scale): usefulness of the P-scale to a rehabilitation services network. *BMC Public Health*, *17*(1), 934. https://doi.org/10.1186/s12889-017-4945-9

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada assessment pendidikan*. Trim Komunikata.

UNICEF. (2023, May). *Adolescents*: *Investing in a safe, healthy and productive transition from childhood to adulthood is critical*. UNICEF. https://data.unicef.org/topic/adolescents/overview/

Werneke, U., Goldberg, D. P., Yalcin, I., & Üstün, B. T. (2000). The stability of the factor structure of the General Health Questionnaire. *Psychological Medicine*, *30*(4), S0033291799002287. https://doi.org/10.1017/S0033291799002287