

Adaptation and Psychometric Validation of the Bangla Quantitative Workload Inventory among Bangladeshi Ready-Made Garment Workers: Evidence from Classical Test Theory and Item Response Theory

Md. Reyad Hossen, Msd. Sanjida Akter Mim, Md. Refayet Ullah, Md. Imam Hossain

University of Chittagong, Bangladesh

DOI: <https://doi.org/10.47772/IJRISS.2026.100400611>

Received: 17 April 2026; Accepted: 22 April 2026; Published: 20 May 2026

ABSTRACT

The quantitative workload is the most crucial factor influencing employee well-being and mental health, especially in Bangladeshi garment workers. This study aimed to culturally adapt and validate the Quantitative Workload Inventory (QWI). The adaptation process followed established cross-cultural guidelines, including forward and backward translation, expert review, and pre-testing. The finalized Bangla QWI was administered to 406 ready-made garment (RMG) workers, and its psychometric properties were evaluated using classical test theory (CTT) and item response theory (IRT) approaches. Findings revealed good internal consistency (Cronbach's $\alpha = .821$, McDonald's $\omega = .834$), satisfactory item discrimination, an acceptable test-retest reliability ($r = .728$), and a single-factor structure of the Bangla-QWI supported by Confirmatory Factor Analysis (CFA). Graded Response Model (GRM) analysis supported item performance. Measurement invariance findings across gender showed configural, metric, and scalar invariance, supporting the scales' comparability between male and female workers. Convergent validity was evidenced by a significant positive association with mental health difficulties ($r = .330, p < .001$). It is a valid and culturally reliable tool for measuring workload among RMG workers. It helps researchers and organizations in designing interventions to manage workload and promote worker health.

Keywords: Adaptation, IRT analysis, Psychometric validation, Quantitative Workload Inventory (QWI), ready-made garment (RMG) workers

INTRODUCTION

Workload is widely recognized as one of the most influential job features affecting workers' performance, well-being, and organizational productivity (Ganster & Rosen, 2013). In modern workplaces, workers often experience high production demands, tight deadlines, and time pressure, which substantially influence their mental and physical health. It is conceptualized as a key job demand that requires sustained efforts and consumes workers' physical and psychological resources. According to the Job Demands-Resources (JD-R) model, excessive job demands like heavy workload can reduce workers' resources and lead to strain, emotional exhaustion, and reduced well-being, particularly when job resources are insufficient to buffer these demands (Bakker & Demerouti, 2007). Excessive workload among RMG workers was the top stressor that directly impacted the experience of burnout and psychological stress, and decreased job performance, especially due to long working hours and unrealistic target production (Salsabila et al., 2021; Wicaksono et al., 2024; Yuan et al., 2022). Consequently, the precise measurement of workload is crucial for understanding its effects on workers' physical and mental health, recognizing occupational risks, and designing effective organizational interventions.

Among the several types of workload, including Quantitative, Qualitative, Cognitive, Emotional, and Physical workload. Quantitative workload is the amount of work a person must complete, often within a time frame. It includes task volume, deadlines, and time constraints (Karasek, 1979; Spector et al., 1998). It was significantly associated with job stress, job dissatisfaction, turnover intention, role conflict, frustration, emotional exhaustion, and reduced job satisfaction among RMG workers (Akhter et al., 2017; Bowling et al., 2015; Spector et al.,

1998). It's interesting to note that role conflict was more strongly associated with depression than with frustration, whereas workload was more powerfully associated with frustration than despair (Spector et al., 1998). Accurate assessment of workers' perception of workload is therefore critical for both research and practice in organizational psychology and occupational health.

The Quantitative Workload Inventory (QWI) developed by Spector et al. (1998) is among the most widely used tools for measuring quantitative workload. It is one of the most widely used self-report tools for measuring the volume of work and speed, with high reliability, strong psychometric properties, a single-factor structure, and widely accepted cross-cultural adaptation, such as Pakistan, Poland, and the USA (Baka & Bazińska, 2016; Rasul & Masood, 2022; Spector et al., 1998). An extensive study has demonstrated its reliability and validity across diverse occupational groups and cultural contexts (Bowling et al., 2015). Cross-cultural adaptations have confirmed its strength; for instance, an Urdu version among Pakistani employees supported the scale's single-factor structure and psychometric soundness (Rasul & Masood, 2022). Additionally, studies conducted in Europe and North America have revealed that the QWI reliably predicts job-related outcomes, including job satisfaction, job performance, and psychological strain (Balducci et al., 2021).

Despite the QWI's international applicability, validated tools for assessing quantitative workload in Bangla are unusual. This limitation is particularly prominent in labor-intensive sectors, which employ millions of workers and contribute substantially to the national economy. Workers in this sector face high production targets, long working hours, repetitive tasks, and strict deadlines, and conditions that may enhance workload and increase the risk of job stress (Akhter et al., 2017; Bakhshi et al., 2019; Mamun et al., 2022). Without culturally adapted and validated tools, it is challenging to systematically evaluate workload and its effect on workers' well-being in this sector.

Cross-cultural adaptation and validation of psychological measures are crucial because tools developed in one language may not retain reliability or validity when used in a different cultural context without rigorous adaptation. Methodological guidelines for scale adaptation emphasize translation and back-translation, evaluation of factor structure, assessment of reliability, and examination of construct validity (Beaton et al., 2000; Brislin, 1970; Hinkin, 1998). Such procedures confirm that tools accurately capture the intended constructs and allow valid comparisons across cultures.

The association between quantitative workload and mental health difficulties can be theoretically explained using the JD-R model. In this theory, workload requires sustained effort and can lead to strain and emotional exhaustion when workers lack sufficient resources to cope with these demands (Bakker et al., 2017). Within labor-intensive sectors, high quantitative workload constitutes a significant job demand that may deplete workers' physical and psychological resources, ultimately affecting well-being, job satisfaction, and turnover intentions. Accurate measurement of this demand is critical to operationalizing the JD-R model and designing effective workplace interventions.

The RMG industry contributes vital role in the economic development of many developing countries, including Bangladesh, Vietnam, China, India, and Turkey. In Bangladesh, the RMG industry of Bangladesh is a major contributor to national economic growth and export earnings. According to the Bangladesh Garment Manufacturers and Exports Association, the industry employs approximately four million workers, which accounts for over 84% of export revenue. In recent years, Bangladesh has emerged as the world's second-largest apparel exporter. Despite its prosperity, the garment sector is largely dependent on unfavorable labor practices such as awful working conditions, inadequate salaries, and long hours. Previous studies have shown that workers in labor-intensive sectors such as manufacturing garments frequently experience high job demands and workload pressures (Akhter et al., 2017). These conditions may increase the risk of job stress and negatively affect workers' well-being.

Although the QWI has been widely validated across several countries and occupational groups in the world, no study has examined its psychometric validation in the Bangla language. The absence of culturally validated tools limits the ability to accurately measure workload in the garment sectors. The current study addresses this gap by adapting and validating the Bangla-QWI among RMG workers. This study contributes to occupational health research in developing countries and provides researchers and practitioners with a reliable tool for measuring

workload in Bangladeshi workplaces.

There are several purposes of this current study. First, we expected to translate and adapt the QWI into the Bangla language. Second, we evaluated the construct validity of the QWI using Confirmatory Factor Analysis. Third, we aimed to measure and compare the internal consistency using McDonald's ω and Cronbach's α . Fourth, we aimed to examine item discrimination through Item Response Theory. Fifth, we anticipated determining measurement invariance across male and female workers. Finally, we expected that quantitative workload would have a significant positive association with mental health issues.

METHODS

Participants

We conducted an on-site survey of 406 RMG workers in Bangladesh with a convenience sampling approach. We gathered data from September 2024 to January 2025 using face-to-face structured interviews. The majority of the participants were female ($n = 254, 53\%$) and married ($n = 298, 73.4\%$), and all participants' ages ranged from 18 to 48 ($M = 26.51$ years, $SD = 6.20$). Most of the respondents had a monthly income ranging from 15100 to 30000 BDT ($n = 259, 63.8\%$) and job tenure below six years ($n = 293, 72.2\%$) (see Table 1).

Table 1 Socio-demographic characteristics of participants.

Demographic Variables	Categories	Number	Percentage (%)
Age	<28 years	250	61.6%
	28-37 years	129	31.8%
	> 37 years	27	6.7%
Sex	Female	215	53%
	Male	191	47%
Tenure	1-5 years	293	72.2%
	6-10 years	67	16.5%
	11-15 years	33	8.1%
	16-20 years	11	2.7%
	> 20 years	2	.5%
Monthly Income	<15100 BDT	129	34.2%
	15100-30000 BDT	259	63.8%
	>30100 BDT	8	2%
Marital Status	Unmarried	107	26.4%
	Married	298	73.4%
	Divorced	1	.2%

Note: BDT = Bangladeshi Taka, Conversion rate in April 2025 (approx.): 1 BDT = 0.0082 USD.

Procedure

Cross-cultural adaptation

The International Test Commission (ITC, 2018) and Sousa and Rojjanasrirat (2011) standards were followed in adapting the QWI to Bangla and Bangladeshi culture through forward, back-translations, synthesis, and pre-

testing. The adapted version was assessed, compared, and synthesized by an expert panel that included an assistant professor of psychology who spent more than five years working in the organizational behavior domain and psychometrics, two students of psychology, and a forward translator. Backward translators were excluded from the first synthesis of the forward translations to reduce the possibility of bias in the translation process. A total of four translators translated every item independently during the forward and backward translation stages. Additionally, during the final review and synthesis stage, the panel of experts evaluated whether every item in the final version, including item replies and instructions relevant, appropriate, and complete, in addition to comparing the back-translated versions with the original QWI scale. In these phases, the expert panel's subjective compromise was taken into consideration. At every level of this process, the linguistic, psychological, and cultural factors were taken into consideration.

Pre-testing

Pre-testing was done on the final synthesized version. Fifty workers took part in the pre-testing phase (this sample size was chosen based on suggestions from Perneger et al. (2015)). The objectives, significance, and ethical details of the current investigation were explained to the participants. Respondents were selected based on their spontaneous enthusiasm for participating in the pre-testing. To make sure the participants understood the meaning of the items, whether they interpreted them similarly to other participants, and whether they thought the items were relevant and thorough for assessing workload in the Bangladeshi culture and language, a cognitive interview was carried out. To evaluate the scale's comprehensibility, they were asked to examine the item's intended meanings, replies to the item, and instructions. Additionally, if they thought any statement was unclear or ambiguous, they were asked to recommend the best suitable word or phrase that is frequently used in the local Bangla language. The expert panel assessed each item's replies (notes were also obtained during cognitive debriefing), including the corrected item-to-total correlations. The items were revised if necessary or changed if there were any differences, inconsistencies, or ambiguities in the participants' verbal expressions on the items' intended meaning. Every item was pertinent and easily comprehended by those respondents, according to the panel's expert assessments of cognitive questioning and measure item-to-total correlations. Next, a convenience sampling technique was used to pick 50 RMG workers as a sample for testing the translated Bangla QWI. The corrected item-total correlations (CITC) ranged from .38 to .50, and Cronbach's α was .80. Throughout this process, we considered linguistic, psychological, and cultural contexts. This scale was incorporated into the final research.

Ethical Approval

To carry out this research, we followed the Declaration of Helsinki (World Medical Association, 2013). We shared an information sheet with each participant that included a brief description of the current research and measures taken to ensure participants' privacy and their right to withdraw from the research at any moment. Upon receiving the verbal and written consent from the participants, we administered the questionnaire. We took permission from the parents for some participants under the age of 18. This study received ethical approval from the Ethics Review Board of the institutions (Ref. No.: AERB-FBSCU-20250527-(1)).

Measures

Quantitative Workload Inventory (QWI)

The QWI was initially intended to evaluate the perceived quantity of work in terms of volume and pace. This scale was developed by Spector et al.(1998) and consisted of five items. Respondents rate the frequency of each statement regarding the quantity of work, ranging from 1 (less than once a month or never) to 5 (many times a day). High scores indicate a greater level of workload. The total score ranged from 5 to 25. It showed satisfactory internal consistency in this study (Cronbach's α = .821, McDonald's ω = .834).

General Health Questionnaire (GHQ-12)

Goldberg et al. (1979) created the original GHQ, which was adapted and translated to measure the participants' mental health problems (Sorcar & Rahman, 1990). It is a self-administered screening test used to find individuals

who could have a diagnosable mental illness. It is focused on two events: distressing experiences and the inability to carry out one's regular health functions. Using a 4-point rating system, each item on the scale asks if the participant has recently noticed a particular behavioral symptom. The responses were given a total of 0, 1, 2, and 3, and the scoring from 0 (always) to 3 (never). There were six negative and positive items out of twelve items. Out of the twelve items, items no. 1, 3, 4, 7, 8, and 12 were reverse-coded. Higher scores indicate greater mental health problems. A total score ranges between 0 and 36. It showed acceptable internal consistency (Cronbach's $\alpha = .782$, McDonald's $\omega = .774$).

Statistical Analysis

To conduct the analyses, SPSS (version 27) and R (version 4.4.3) were used. SPSS was used for the following purposes: data cleaning, descriptive analyses of the sample, CITC, inter-item correlation, Cronbach's alpha, omega, test-retest reliability, and Pearson product-moment correlation. R was utilized for the following purposes: IRT analysis, CFA, multigroup CFA, factor loading, and other reliability measures. Descriptive analysis includes the mean and the standard deviation. We checked the normality assumption using skewness and kurtosis [a value between -2 and +2 suggests normality (Gravetter & Wallnau, 2014)]. The item's ability to discriminate between different levels of latent traits was assessed by corrected item-total correlation (CITC). An acceptable degree of reliability is generally defined as 0.6 to 0.7, while a very good level is defined as 0.8 or higher. Nevertheless, values greater than 0.95 are not always favorable because they may indicate redundancy (Hulin et al., 2001).

The classical test theory (CTT) was utilized to evaluate the Bangla QWI psychometric properties. We evaluated the following: composite reliability (accepted values ≥ 0.70), internal consistency reliabilities (Cronbach's alpha, McDonald's omega, and using the Spearman-Brown formula; accepted value ≥ 0.70), average variance extracted (AVE) (accepted values ≥ 0.50), inter-item correlations ideal range 0.15 to 0.50 but 0.80 are still considered acceptable only for unidimensional or short scales (Briggs & Cheek, 1986) and CITC (minimum of 0.30; DeVellis, 2021). A good measurement model is demonstrated by the variables' AVE being over 0.5 (Basheer et al., 2019). Additionally, $\geq .30$ should be the prominent factor loading (Field, 2024). Test-retest reliability over the two-week interval was examined using Pearson correlation coefficients and intraclass correlation coefficients between Time 1 and Time 2 scores for the 50 participants.

The CFA was completed to test the single-factor structure of the Bangla-QWI. Model fits were evaluated using the χ^2/df value, goodness of-fit index (GFI), comparative fit index (CFI), Tucker-Lewis index (TLI), standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA). A good fit is specified by a χ^2/df ratio value of ≤ 2 (Schermelleh-Engel et al., 2023). However, this value should be less than 5 (Marsh et al., 1998). A good fit model is specified by GFI, CFI, and TLI values $\geq .95$ (Hu et al., 1999). Additionally, a better-fitted model is suggested by an SRMR value $\leq .08$ and an RMSEA value between .06 and 0.08 (Schreiber et al., 2006).

Under the IRT, we conducted item analysis utilizing the graded response model (GRM). The performance of individual items about the underlying qualities is provided by IRT models. The item involvement across the trait distribution can be estimated with the help of these models. Before using the GRM, the assumptions of unidimensionality and monotonicity were examined. A set of items is considered Unidimensional if it evaluates just a solitary trait (Strout, 1990). Loevinger's H coefficients and CCFA model fits can be used to evaluate unidimensionality. Standard guidelines state that Loevinger's H values between .30 - .39 indicate weak, .40 - .49 moderate, and $\geq .50$ high unidimensionality (Sijtsma & Molenaar, 2002). R package Mokken version 3.0.2 was used to evaluate Loevinger's H coefficients and monotonicity. GRM analysis was conducted using this R package version to evaluate the Bangla QWI item qualities. Two parameters were assessed in this model: the discrimination parameter [a] or slope coefficient, and the threshold coefficient [b]. We evaluated item slope parameter (a) values in accordance with Baker's (2001) recommendation, which states that values .35 to .64 indicate minor discrimination power in the items in differentiating between low and high attribute levels, values .65 to 1.34 indicate moderate discrimination power, values 1.35 to 1.69 indicate high discrimination, and values ≥ 1.70 indicate very high discrimination. The value of [b] shows the degree of fundamental trait at which a person has a 50% possibility of recommending a category. Monotonicity implies that the possibility of supporting an answer to an item does not decrease as the latent trait increases (Linden et al., 1997). Both

statistical and graphical approaches were used to evaluate the items' monotonicity. Using the crit statistics, the significance of monotonicity desecration was evaluated. According to Sijtsma and Molenaar (2002), a crit value < 40 was regarded as not constituting a violation of monotonicity. Additionally, the Rho coefficient [$>.70$ acceptable; Sijtsma & Molenaar, 2002] for each scale was evaluated. In addition to these constraints, test information curves (TIFs) and item characteristic curves (ICCs) for this scale were assessed.

Local independence implies that there is no statistical association between an individual's response to several items (McDonald, 1981). Some authors claimed that if unidimensionality is true, local autonomy is obvious (Hays et al., 2000). Yen's Q3 residual coefficient (Yen, 1984) was utilized in this study to measure the local independence of items. The two items may have local independence if the Q3 coefficient is more than .20 (Chen et al., 1997; Tate, 2003).

Next, A series of multigroup CFA models was then conducted to assess the QWI Bangla's measurement invariance across sociodemographic categories. In the current study, measurement invariance was measured for both males and females. We employed ΔCFI , $\Delta RMSEA$, and $\Delta SRMR$ to determine invariance in this study, as the $\Delta\chi^2$ is an extremely stringent (MacCallum et al., 1992) and sensitive to sample size (Dimitrov, 2010). A proof of measurement invariance, Chen (2007) recommended $\Delta CFI \geq 0.010$, $\Delta SRMR \leq 0.01$, and $\Delta RMSEA \leq 0.015$ to demonstrate measurement invariance. In addition to the statistical tests mentioned above, the validity of the QWI was evaluated using the Pearson product-moment correlation coefficient.

RESULTS

Basic Descriptive Statistics

Table 2 shows the mean (M) and standard deviation (SD), with skewness and kurtosis, CITC, and factor loading of each item of QWI. The value appeared to be a normal distribution, as indicated by the skewness (ranging from -.546 to 1.416) and kurtosis (ranging from -1.387 to .422) values. Item analysis findings demonstrated that every item had acceptable CITC (varying between .483 and .669). Inter-item correlations given all items are significantly associated with each other (ranging from .344 to .612) (see Table 2).

Table 2 Item-level psychometric properties of the Quantitative Workload Inventory (QWI).

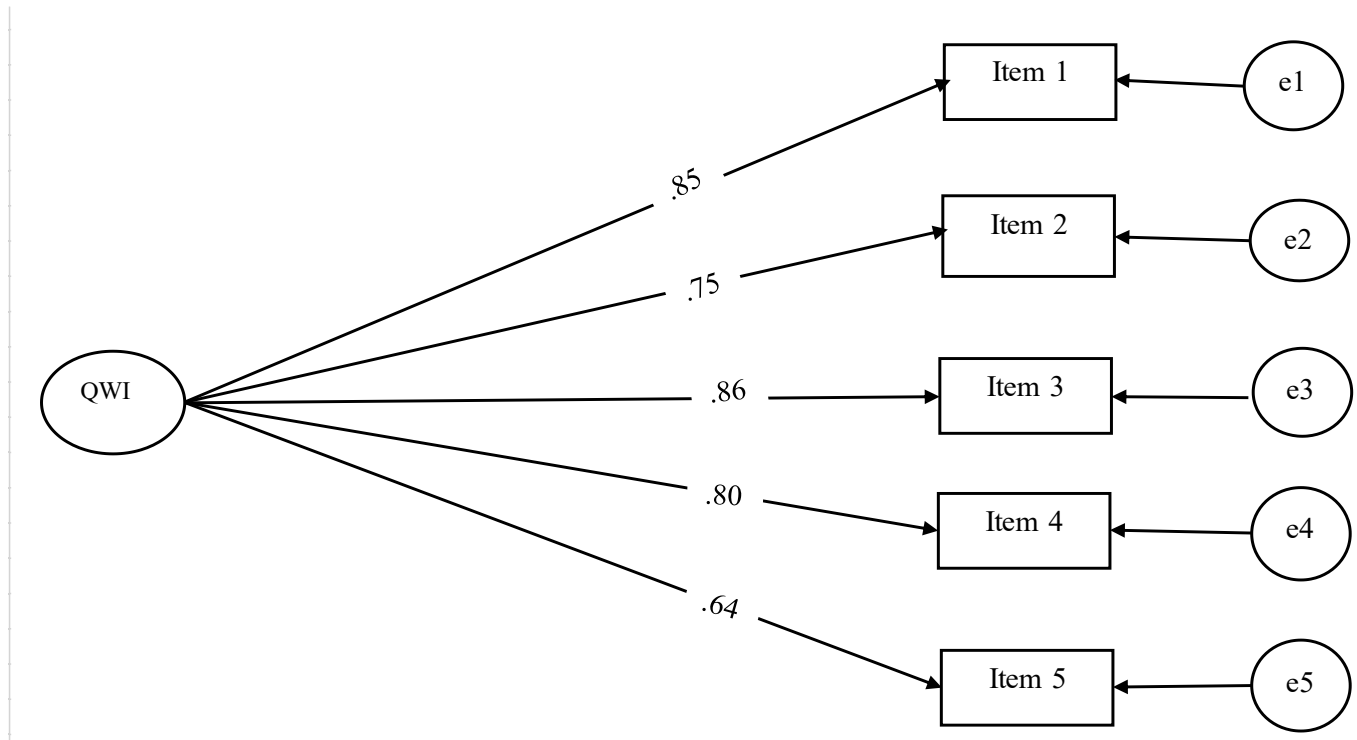
Items	M	SD	Skewness	Kurtosis	CITC	Factor loading
Item 1	1.85	1.402	1.416	.422	.645	.854
Item 2	2.57	1.552	.492	-1.347	.632	.752
Item 3	2.21	1.586	.868	-.950	.669	.859
Item 4	2.33	1.528	.743	-1.015	.656	.804
Item 5	3.52	1.619	-.546	-1.387	.483	.639

Note: M = Mean; SD = Standard deviation; CITC = corrected item-total correlation.

CFA-based Construct Validity

We investigated a single-factor model for Bangla QWI, as recommended by the original authors (Spector et al., 1998) and earlier validation studies in Pakistan (Rasul & Masood, 2022). According to the CFA findings: $\chi^2(5) = 9.028$, $p = 0.108$; TLI = .996; CFI = .998; GFI = .999; RMSEA = .045, 95% confidence interval [CI] [.029, .049]; SRMR = .025. All items of the Bangla QWI were significantly loaded, with factor loadings (λ) ranging from .64 to .85 (see Table 3, Figure 1).

Figure 1 Single-factor model of the Bangla-Quantitative Workload Inventory (QWI).



These results suggest that the Bangla-QWI had a single-factor structure.

IRT-based GRM

Discrimination (a) and threshold parameters (b1-b4) for every item are among the item parameters resulting from the Graded Response Model (GRM) and are shown in Table 4. The majority of items can efficiently distinguish between people across the latent characteristic spectrum, indicated by the discrimination values, which ranged from 1.331 to 2.768. Item 5 had the slope coefficient (a = 1.331), which indicates moderate discrimination.

Table 4 Item Response Theory (IRT) Graded Response Parameters for QWI.

Items	a	b1	b2	b3	b4
Item 1	2.768	0.449	0.934	1.078	1.420
Item 2	1.972	-0.521	0.378	0.540	1.150
Item 3	2.766	0.139	0.609	0.739	1.107
Item 4	2.446	-0.154	0.496	0.759	1.186
Item 5	1.331	-1.339	-0.466	-0.466	0.267

Note: a = discrimination parameter; b1-b4 = threshold parameters for item response categories.

The trait levels at which participants are likely to support a response in the following category or above are represented by the threshold values (b1-b4).

Table 5 revealed that the range of Loewinger's H coefficients was from .545 to .612. It indicates that all items had strong scalability. These results indicate that the items provide a significant contribution to a unidimensional scale. The ranges of Q3 residual coefficients (Yen's Q3) were -0.033 to -0.330, all of which are much below the .20 threshold of concern. This suggests that there was no problematic local dependence between the items. Monotonicity assumptions were satisfied for all items, with no violation detected.

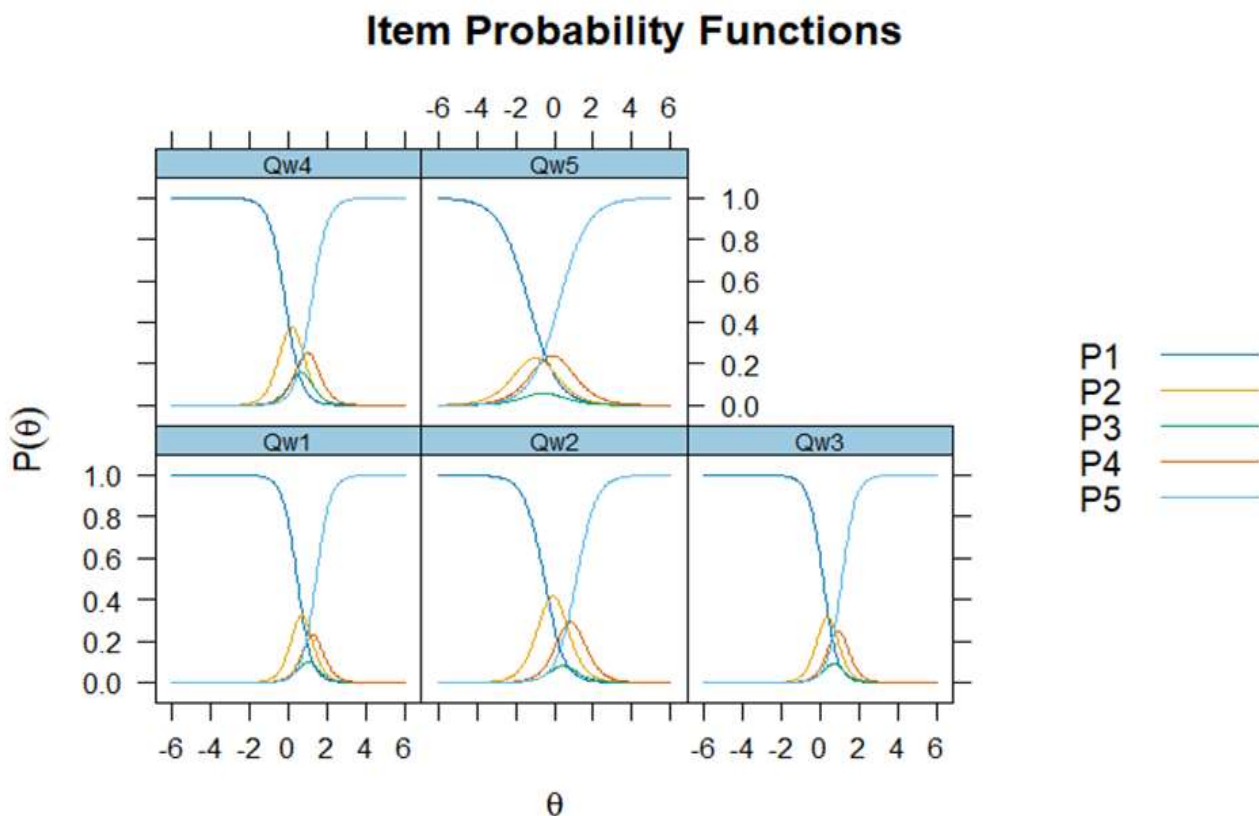
Table 5 H coefficient, Q3 coefficient, and monotonicity outputs of the Bangla- QWI.

Items	H	Monotonicity				Residual coefficients (Yen's Q3 Coefficients)				
		#ac	#vi	#zsig	crit	Item1	Item2	Item3	Item4	Item5
Item1	.612	17	0	0	0	1				
Item2	.545	24	0	0	0	-.207	1			
Iem3	.585	24	0	0	0	-.229	-.292	1		
Iem4	.564	24	0	0	0	-.278	-.157	-.330	1	
Item5	.551	24	0	0	0	-.115	-.033	-.089	-.103	1

Note: #ac = (active comparison) #vi = (violations) #zsig = (significant violation), crit = critical ratio; H = Levinger’s H coefficients.

Item probability curves revealed properly ordered response categories across the latent workload trait (θ), indicating appropriate category functioning (see Figure 2).

Figure 2 Item Probability Function of Bangla-Quantitative Workload Inventory (QWI).

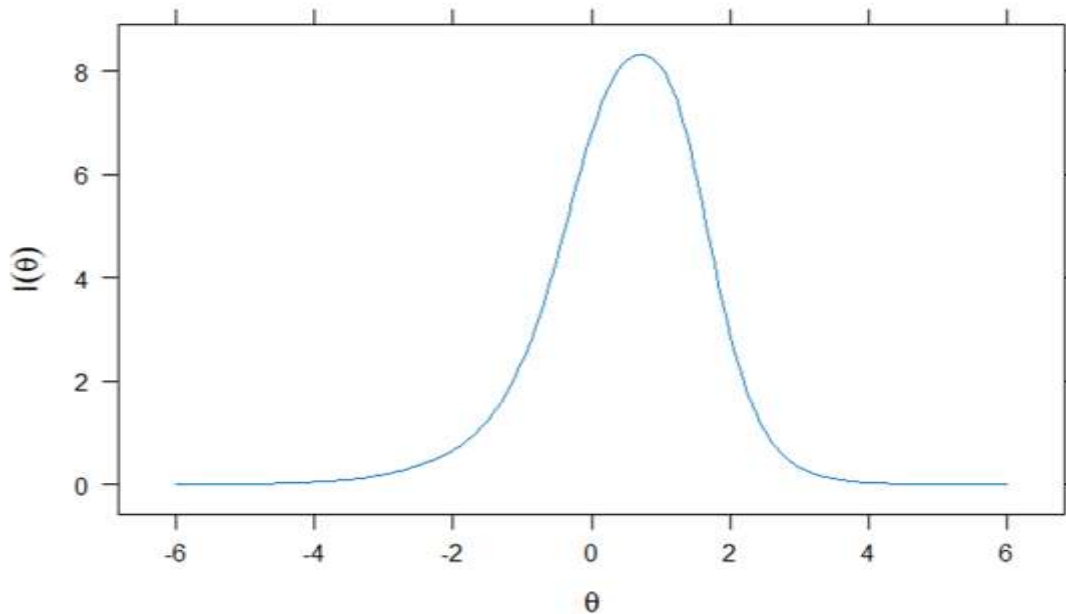


Note: X-axis (θ) = latent workload level; Y-axis $I(\theta)$ = probability of endorsing a given response category at each θ level.; (P1-P5) = P1 for the lowest workload response and P5 for the highest response.

The items showed greater sensitivity at moderate to high workload levels, with crossover points primarily occurring in higher response categories (p1-p5). The test information curve indicated that measurement precision peaked around $\theta \approx .5$ (see Figure 3).

Figure 3 Scale Information Curve of Bangla-Quantitative Workload Inventory (QWI).

Scale Information Curve for Quantitative Workload Inventory



Note: The X-axis (θ) = latent trait level (workload); the Y-axis $I(\theta)$ = test information curve

CTT-based Measurement invariance across gender

The findings of multigroup CFA demonstrated that the QWI-Bangla was configural, metric, and scalar invariant across gender. The configural invariance model of gender had a satisfactory model fit ($\chi^2 = 11.179$, $df = 10$, p value = .344, CFI = .999, RMSEA = .024, SRMR = .064). The CFI and RMSEA values for the configural vs metric models differed slightly (-.001, .024, respectively), metric vs scalar models (.000, .000, respectively). These findings accepted full measurement invariance, suggesting that the QWI measures workload equivalently in both males and females.

Reliability and Validity

The reliability of the Bangla QWI’s internal consistency was evaluated utilizing Cronbach’s alpha (α) and McDonald’s omega (ω). The QWI demonstrated good internal consistency in indices in Bangladeshi culture, with Cronbach’s alpha (α) of .821 and McDonald’s omega (ω) of .834 (see Table 3).

Table 3 Psychometric properties of the Bangla- Quantitative Workload Inventory (QWI) at the scale level.

Psychometric properties	Value	Suggested cutoff
Cronbach’s alpha	.821	≥ 0.70
McDonald’s omega	.834	≥ 0.70
Test-retest reliability	.728	≥ 0.70
AVE	.617	≥ 0.50
Mean inter-item correlation	.482	$.15 \leq r \leq .50$
Composite Reliability	.921	≥ 0.70
Split-half reliability	.932	≥ 0.70
Rho coefficient	.913	≥ 0.70

Ferguson's Delta	.911	≥ 0.70
SEM	.650	$< SD/2$

Note: AVE, average variance extracted; CR, composite reliability; SD, standard deviation; df, degrees of freedom; SEM, Standard error of measurement.

The test-retest reliability at a two-week interval demonstrated acceptable consistency ($r = .73$). The composite reliability ($r = .921$), the Rho coefficient (.91), split-half reliability (.93), and Ferguson's delta (.91) reflected strong internal consistency and adequate discriminatory capacity. Quantitative workload was significantly positively associated with mental health issues ($r = .330, p < .001$) by correlation analyses.

DISCUSSION

The current study evaluated the psychometric properties of the Bangla Quantitative Workload Inventory (QWI) among Bangladeshi RMG workers using both Classical Test Theory and Item Response Theory. Overall, the findings provide strong evidence supporting the reliability, validity, and structural integrity of the adapted tool within a labor-intensive sector in Bangladesh. At the item level, all QWI-Bangla items demonstrated satisfactory corrected item-total correlations, indicating that each item contributes meaningfully to the overall construct of quantity of workload. Inter-item correlations were also within the acceptable range, suggesting adequate homogeneity without redundancy. These findings are consistent with prior cross-cultural validations of the QWI, which have reported similar consistency patterns across diverse occupational settings (Baka & Bazińska, 2016; Rasul & Masood, 2022).

At the scale level, multiple reliability indices, including composite reliability, Cronbach's alpha, and McDonald's omega, indicated good internal consistency. The test-retest reliability further revealed temporal stability, suggesting the quantity of workload is a relatively stable construct over short periods. These findings align with previous international validation studies of QWI (Baka & Bazińska, 2016; Spector et al., 1998).

The confirmatory factor analysis (CFA) supported a unidimensional structure of the QWI-Bangla, consistent with the original conceptualization (Spector et al., 1998) and subsequent cross-cultural validations (Baka & Bazińska, 2016; Rasul & Masood, 2022). The model showed excellent fit indices, with all items loading significantly on a single latent factor. This supports the theoretical assumption that quantitative workload represents a unified construct reflecting the amount and pace of work demands. Notably, this structure is consistent with the Job Demands-Resources (JD-R) model, which conceptualizes workload as a key job demand that may lead to strain when job resources are insufficient (Bakker & Demerouti, 2007).

The IRT analyses provided further insights into item functioning. The graded response model (GRM) findings indicated that most items exhibited high discrimination parameters, recommending effective differentiation across levels of the latent workload trait (Baker, 2001). Threshold parameters followed an ordered pattern, confirming appropriate response category functioning. The test information function indicated that the inventory provides the highest measurement precision at moderate to high levels of workload, with lower precision at the lower end of the continuum. This suggests that the Bangla-QWI is particularly sensitive in identifying individuals experiencing higher workload demands, which is especially relevant in high-pressure work environments.

Additionally, Mokken scale analysis indicated strong scalability, supporting the single-factor structure. The absence of problematic local dependence, as indicated by acceptable Q3 residual values, further confirmed the independence of the item responses. The monotonicity assumption was also satisfied, reinforcing the appropriateness of applying IRT models.

Measurement invariance analysis using multigroup CFA demonstrated configural, metric, and scalar invariance across gender, indicating that the scale operates equivalently for both male and female workers. This allows for meaningful comparisons of workload across groups and is aligned with previous studies (Baka & Bazińska, 2016).

Furthermore, convergent validity was further supported through correlation analysis. Quantitative workload was significantly positively associated with mental health issues, consistent with theoretical expectations and prior empirical findings linking high workload to psychological distress (Bakker & Demerouti, 2007).

Practical Implication

This validated Bangla-QWI provides a reliable and culturally valid tool for measuring the quantity of workload in the RMG sector. From a practical perspective, this instrument enables organizations to systematically identify employees and work units exposed to heavy workload demands, which are known risk factors for adverse outcomes like psychological distress, reduced job satisfaction, and increased turnover intentions. In line with the Job Demands-Resources (JD-R) model, organizations can use the QWI to monitor job demands and design targeted interventions, like workload redistribution, shift adjustments, and improve task planning. This scale can also be incorporated into routine organizational assessments to measure the effectiveness of workload management policies over time. Furthermore, practitioners and policymakers may use this tool to generate evidence for developing labor regulations and workplace standards aimed at protecting workers' well-being in labor-intensive sectors

Limitations and Future Research Directions

Despite the encouraging results, several limitations should be considered. First, the current data were gathered from garment workers in two districts using a convenience sampling technique, which may limit generalizability. Future studies may consider a nationally representative sample across diverse groups of employees, including nurses, institutions, NGOs, and both the private and public sectors. Second, self-report measures were used to collect the data, which may introduce response bias, including social desirability bias. Future research could integrate behavioral or objective measures to complement self-reports. This study didn't integrate discriminant validity. Future research could check discriminant validity. Finally, the cross-sectional design excludes examinations of predictive validity; longitudinal studies are needed to measure how quantitative workload predicts outcomes such as mental health, turnover intention, and job performance. Addressing these limitations will strengthen the scale's applicability and broader understanding of workload dynamics in Bangladeshi workplaces.

CONCLUSION

The Bangla Quantitative Workload Inventory (QWI) demonstrated strong psychometric properties, including good internal consistency reliability, strong item discrimination, and a single-factor structure consistent with international validation. The scale reliably assesses the quantity of workload among Bangladeshi RMG workers, offering a culturally adapted tool aligned with the Job Demands- Resources (JD-R) model. It enables practitioners, researchers, and policymakers to identify high demands work-environments and examine their effects on worker well-being, mental health, job satisfaction, and turnover intentions. Additionally, the Bangla-QWI supports cross-cultural comparisons and longitudinal studies, providing insights into how workload interrelates with other occupational stressors and job resources. Practically, it can guide evidence-based interventions, workload management strategies, and initiatives to promote psychological safety and improving but individual and organizational outcomes. Thus, the Bangla-QWI is a valid, reliable, contextually relevant tool for advancing occupational health research and practice in labor-intensive sectors.

REFERENCES

1. Akhter, S., Rutherford, S., Akhter Kumkum, F., Bromwich, D., Anwar, I., Rahman, A., & Chu, C. (2017). Work, gender roles, and health: neglected mental health issues among female workers in the ready-made garment industry in Bangladesh. Taylor & FrancisS Akhter, S Rutherford, F Akhter Kumkum, D Bromwich, I Anwar, A Rahman, C ChuInternational Journal of Women's Health, 2017•Taylor & Francis, 9, 571–579. <https://doi.org/10.2147/IJWH.S137250>
2. Baka, Ł., & Bazińska, R. (2016). Polish adaptation of three self-report measures of job stressors: The interpersonal conflict at work scale, the quantitative workload inventory and the organizational constraints scale. *International Journal of Occupational Safety and Ergonomics*, 22(1), 32–39.

- <https://doi.org/10.1080/10803548.2015.1116816>
3. Baker, F. B. (2001). *The Basics of Item Response Theory*. Second Edition. Eric.Ed.Gov, 1–187.
 4. Bakhshi, E., Mazloumi, A., Hoseini, S. M., Bakhshi, E., Mazloumi, A., Hoseini, S. M., Bakhshi, E., Mazloumi, A., Hoseini, S. M., Bakhshi, E., Mazloumi, A., & Hoseini, S. M. (2019). Relationship Between Mental Fatigue and Mental Workload Among Nurses. *Zahedan Journal of Research in Medical Sciences* 2019 21:1, 21(1), e83082. <https://doi.org/10.5812/ZJRMS.83082>
 5. Bakker, A. B., & Demerouti, E. (2007). The Job Demands-Resources model: State of the art. *Journal of Managerial Psychology*, 22(3), 309–328. <https://doi.org/10.1108/02683940710733115/FULL/HTML>
 6. Bakker, A., health, E. D.-J. of occupational, & 2017, undefined. (2017). Job demands–resources theory: Taking stock and looking forward. *Psycnet.Apa.Org* AB Bakker, E Demerouti *Journal of Occupational Health Psychology*, 2017•*psycnet.Apa.Org*, 22(3), 273–285. <https://doi.org/10.1037/ocp0000056>
 7. Balducci, C., Alessandri, G., Zaniboni, S., Avanzi, L., Borgogni, L., & Fraccaroli, F. (2021). The impact of workaholism on day-level workload and emotional exhaustion, and on longer-term job performance. *Taylor & Francis*, 35(1), 6–26. <https://doi.org/10.1080/02678373.2020.1735569>
 8. Basheer, M., Hafeez, M., Hassan, S., Paradigms, U. H.-, & 2018, undefined. (n.d.). Exploring the role of TQM and supply chain practices for firm supply performance in the presence of organizational learning capabilities: a case of textile firms in. *Researchgate.Net* MF Basheer, MH Hafeez, SG Hassan, U Haroon *Paradigms*, 2018•*researchgate.Net*.
 9. Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Journals.Lww.Com* DE Beaton, C Bombardier, F Guillemin, MB Ferraz *Spine*, 2000•*journals.Lww.Com*, 25(24), 3186–3191. <https://doi.org/10.1097/00007632-200012150-00014>
 10. Bowling, N. A., Alarcon, G. M., Bragg, C. B., & Hartman, M. J. (2015). A meta-analytic examination of the potential correlates and consequences of workload. *Taylor & Francis* NA Bowling, GM Alarcon, CB Bragg, MJ Hartman *Work & Stress*, 2015•*Taylor & Francis*, 29(2), 95–113. <https://doi.org/10.1080/02678373.2015.1033037>
 11. Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Wiley Online Library* SR Briggs, JM Cheek *Journal of Personality*, 1986•*Wiley Online Library*, 54(1), 106–148. <https://doi.org/10.1111/J.1467-6494.1986.TB00391.X>
 12. Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1(3), 185–216. <https://doi.org/10.1177/135910457000100301>
 13. Chen, W., and, D. T.-J. of E., & 1997, undefined. (1997). Local dependence indexes for item pairs using item response theory. *Journals.Sagepub.Com* WH Chen, D Thissen *Journal of Educational and Behavioral Statistics*, 1997•*journals.Sagepub.Com*, 22(3), 265–289. <https://doi.org/10.3102/10769986022003265>
 14. Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling: a multidisciplinary journal*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>
 15. DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.
 16. Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Taylor & Francis* ED Diener, RA Emmons, RJ Larsen, S Griffin *Journal of Personality Assessment*, 1985•*Taylor & Francis*, 49(1), 71–75. https://doi.org/10.1207/S15327752JPA4901_13
 17. Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and evaluation in Counseling and Development*, 43(2), 121-149. <https://doi.org/10.11177/0748175610373459>
 18. Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage publications limited.
 19. Ganster, D. C., & Rosen, C. C. (2013). Work stress and employee health: A multidisciplinary review. *Journals.Sagepub.Com* DC Ganster, CC Rosen *Journal of Management*, 2013•*journals.Sagepub.Com*, 39(5), 1085–1122. <https://doi.org/10.1177/0149206313475815>
 20. Goldberg, D., medicine, V. H.-P., & 1979, undefined. (1979). A scaled version of the General Health Questionnaire. *Cambridge.Org* DP Goldberg, VF Hillier *Psychological Medicine*, 1979•*cambridge.Org*, 9, 139–145. <https://www.cambridge.org/core/journals/psychological-medicine/article/scaled-version-of-the-general-healthquestionnaire/855845C6B366DB4AEDD66C350320D2A9>
 21. Gravetter, F. J., & Wallnau, L. B. (2014). Introduction to the t statistic. *Essentials of statistics for the behavioral sciences*, 8(252).

22. Hays, R., Morales, L., care, S. R.-M., & 2000, undefined. (n.d.). Item response theory and health outcomes measurement in the 21st century. *Journals.Lww.ComRD Hays, LS Morales, SP ReiseMedical Care, 2000*•journals.Lww.Com. <https://doi.org/10.1097/00005650-200009002>
23. Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational research methods, 1*(1), 104-121. <https://doi.org/10.1177/109442819800100106>
24. Hu, L., a, P. B.-S. equation modeling:, & 1999, undefined. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Taylor & FrancisL Hu, PM BentlerStructural Equation Modeling: A Multidisciplinary Journal, 1999*•Taylor & Francis, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
25. Hulin, C. (2001). Can a reliability coefficient be too high?. *J Consum Psychol, 10*(1–2), 55. <https://doi.org/10.1207/S15327663JCP1001>
26. International Test Commission. (2018). ITC guidelines... - Google Scholar. (n.d.).
27. Jama, W. M. A.-, & 2013, undefined. (n.d.). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jamanetwork.Com. Retrieved January 10, 2026, from https://jamanetwork.com/journals/jama/fullarticle/1760318/*
28. Karasek Jr, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative science quarterly, 28*5-308.
29. Linden, W. Van Der, response, R. H.-H. of modern item, & 1997, undefined. (1997). Item response theory: Brief history, common models, and extensions. *Springer, 1–28. https://doi.org/10.1007/978-1-4757-2691-6_1*
30. MacCallum, R., ... M. B.-P., & 1996, undefined. (n.d.). Power analysis and determination of sample size for covariance structure modeling. *Psycnet.Apa.OrgRC MacCallum, MW Browne, HM SugawaraPsychological Methods, 1996*•psycnet.Apa.Org.
31. Mamun, M. A., Hossain, M. S., & Griffiths, M. D. (2022). Mental health problems and associated predictors among Bangladeshi students. *SpringerMA Mamun, MS Hossain, MD GriffithsInternational Journal of Mental Health and Addiction, 2022*•Springer, 20(2), 657–671. <https://doi.org/10.1007/S11469-019-00144-8>
32. McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of mathematical and statistical Psychology, 34*(1), 100-117. <https://doi.org/10.1111/J.2044-8317.1981.TB00621>
33. Marsh, H., bulletin, D. H.-P., & 1985, undefined. (n.d.). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. *Psycnet.Apa.OrgHW Marsh, D HocevarPsychological Bulletin, 1985*•psycnet.Apa.Org.
34. Perneger, T. V., Courvoisier, D. S., Hudelson, P. M., & Gayet-Ageron, A. (2015). Sample size for pre-tests of questionnaires. *SpringerTV Perneger, DS Courvoisier, PM Hudelson, A Gayet-AgeronQuality of Life Research, 2015*•Springer, 24(1), 147–151. <https://doi.org/10.1007/S11136-014-0752-2>
35. Rasul, F., & Masood, S. (2022). Reliability and Validity Estimation of Job Stressor Measures Interpersonal Conflict at Work Scale, Quantitative Workload Inventory and Organizational Constraints Scale in Urdu Language. *Foundation University Journal of Psychology, 6*(1). <https://doi.org/10.33897/fujp.v6i1.354>
36. Salsabilla, A., Setiawan, M., & Juwita, H. A. J. (2022). The effect of workload and job stress on job satisfaction mediated by work motivation. *International Journal of Research in Business & Social Science, 11*(9).
37. Schermelleh-Engel, K., ... H. M.-... psychological research, & 2003, undefined. (n.d.). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Stats.Ox.Ac.Uk.*
38. Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Taylor & FrancisJB Schreiber, A Nora, FK Stage, EA Barlow, J KingThe Journal of Educational Research, 2006*•Taylor & Francis, 99(6), 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>
39. Sijtsma, K., & Molenaar, I. (2002). Introduction to nonparametric item response theory. [https://books.google.com/books?hl=en&lr=&id=xxJaRuLC_4sC&oi=fnd&pg=PP8&dq=Sijtsma,+K.,+%26+Molenaar,+I.+W.+\(2002\).+Introduction+to+nonparametric+item+response+theory.+SAGE+Publications.&ots=votqjCzku6&sig=XOHJz13kc4VsnS7p5D1XtXRrBGw](https://books.google.com/books?hl=en&lr=&id=xxJaRuLC_4sC&oi=fnd&pg=PP8&dq=Sijtsma,+K.,+%26+Molenaar,+I.+W.+(2002).+Introduction+to+nonparametric+item+response+theory.+SAGE+Publications.&ots=votqjCzku6&sig=XOHJz13kc4VsnS7p5D1XtXRrBGw)

40. Sorcar, N. R., & Rahman, A. (1990). Occupational stress, marital status and job satisfaction of working women. *The Dhaka University Studies*, 11(1), 55-61.
41. Sousa RN, V. D., Rojjanasrirat RNC IBCLC, W., & Valmi Sousa, C. D. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *Wiley Online Library* VD Sousa, W Rojjanasrirat *Journal of Evaluation in Clinical Practice*, 2011•Wiley Online Library, 17(2), 268–274. <https://doi.org/10.1111/J.1365-2753.2010.01434.X>
42. Spector, P., psychology, S. J.-J. of occupational health, & 1998, undefined. (1998a). Development of four self-report measures of job stressors and strain: interpersonal conflict at work scale, organizational constraints scale, quantitative workload. *Psycnet.Apa.Org* PE Spector, SM Jex *Journal of Occupational Health Psychology*, 1998•psycnet.Apa.Org, 3(4), 356–367. <https://doi.org/10.1037/1076-8998.3.4.356>
43. Spector, P., psychology, S. J.-J. of occupational health, & 1998, undefined. (1998b). Development of four self-report measures of job stressors and strain: interpersonal conflict at work scale, organizational constraints scale, quantitative workload. *Psycnet.Apa.Org* PE Spector, SM Jex *Journal of Occupational Health Psychology*, 1998•psycnet.Apa.Org, 3(4), 356–367. <https://doi.org/10.1037/1076-8998.3.4.356>
44. Strout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55(2), 293-325.
45. Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied psychological measurement*, 27(3), 159-203. <https://doi.org/10.1177/0146621602250510>
46. Wicaksono, W., Tangibali, C., ... A. U.-P. J. of L., & 2024, undefined. (n.d.). The Happiness Factor: Job Satisfaction as a Buffer between Workload-Included Stress and Productivity. *Pjlss.Edu.Pk*, 12(1), 54–62. <https://doi.org/10.21776/ub.jemis.2024.012.01.6>
47. Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8(2), 125-145. <https://doi.org/10.1177/014662168400800201>
48. Yuan, D., Gazi, M. A. I., Rahman, M. A., Dhar, B. K., & Rahaman, M. A. (2022). Occupational stress and health risk of employees working in the garments sector of Bangladesh: An empirical study. *Frontiers in Public Health*, 10. <https://doi.org/10.3389/fpubh.2022.938248>