Kesmas

Volume 20
Issue 4 *November 2025*

Article 4

11-28-2025

Burnout Assessment Tool Development and Validation for Healthcare Workers: A Psychometric Approach Using Rasch Modelling and Exploratory Factor Analysis

Rosnah Ismail
Universiti Kebangsaan Malaysia, Bangi, drrose@ukm.edu.my

Suriya Kumareswaran Vallasamy
Universiti Kebangsaan Malaysia, Bangi, suriya_kumareswaran@hotmail.com

Recommended Citation

Ismail R , Vallasamy SK . Burnout Assessment Tool Development and Validation for Healthcare Workers: A Psychometric Approach Using Rasch Modelling and Exploratory Factor Analysis. *Kesmas*. 2025; 20(4): 279-287
DOI: 10.7454/kesmas.v20i4.2291
Available at: https://scholarhub.ui.ac.id/kesmas/vol20/iss4/4

Kesmas: Jurnal Kesehatan Masyarakat Nasional (National Public Health Journal)

Burnout Assessment Tool Development and Validation for Healthcare Workers: A Psychometric Approach Using Rasch Modelling and Exploratory Factor Analysis

Rosnah Ismail*, Suriya Kumareswaran Vallasamy

Department of Public Health Medicine, Faculty of Medicine, Universiti Kebangsaan Malaysia, Bangi, Malaysia

Abstract

Burnout remains a significant occupational threat to healthcare workers (HCWs). Existing instruments, e.g., the Maslach Burnout Inventory, have limited applicability in healthcare-specific contexts. This study aimed to develop and validate the Burnout Scale for Assessing Stress in Healthcare Workers (BSAS-HCW), designed to measure cognitive and emotional exhaustion. A cross-sectional survey was conducted among 245 Malaysian HCWs representing diverse roles, including staff nurses (57.7%), house officers (14.3%), medical officers (13.5%), specialists (5.7%), pharmacists (4.5%), and other HCWs. Exploratory factor analysis revealed a two-factor structure with strong internal consistency (Cronbach's alpha: cognitive = 0.95; emotional = 0.97), and all items loaded above 0.4. Rasch analysis supported unidimensionality, robust item fit, and minimal differential item functioning across sex and ethnicity. Person and item reliability exceeded 0.93, with Wright Maps indicating effective item-person targeting. The instrument explained over 70% of the total variance, demonstrating strong psychometric validity. The BSAS-HCW provided a culturally and occupationally relevant tool for assessing burnout among HCWs in high-stress clinical environments. Its dual-domain design addresses the cognitive and emotional toll of burnout more comprehensively than existing tools, enabling more effective detection, monitoring, and intervention strategies across diverse healthcare roles.

Keywords: burnout, exploratory factor analysis, healthcare workers, psychometric validation, Rasch model

Introduction

The healthcare sector is characterized by high demands, with workers facing an increasingly broad range of roles and responsibilities, often exacerbated by resource shortages and uneven distribution.[1] Furthermore, healthcare workers (HCWs) are tasked with managing patients with more complex, chronic, and severe conditions while enduring challenging work environments. These include excessive workloads, frequent workflow disruptions, intense time pressures, a culture of zero tolerance for errors, and minimal social support.[2] Such a condition has likely contributed to the widespread occurrence of burnout and psychological distress among HCWs.

Burnout refers to a state of profound fatigue resulting from prolonged exposure to stress, as outlined in the literature.[3] Individuals suffering from burnout often display severe physical and emotional distress, develop a cynical or indifferent attitude toward their work, and experience a marked decline in professional effectiveness. The latest edition of the International Classification of Diseases-11 specifically identifies burnout as a workplace phenomenon, focusing on these defining syndromes rather than attributing it to broader life circumstances.[4]

Burnout among HCWs remains a significant global issue, with studies reporting a prevalence of 11.23% among nurses in 49 countries[5] and an estimated 37% among doctors worldwide.[6] In Southeast Asia, a pooled analysis indicated a regional burnout prevalence of 17.1% among gastroenterologists, with inter-country variations showing rates exceeding 30% in Malaysia, Singapore, and Brunei Darussalam.[7,8] Although the risk factors for burnout vary by region, a consistent trend is observed: HCWs in the emergency department and intensive care unit are at a heightened risk of burnout.[9,10]

***Correspondence*:** *Rosnah Ismail, Department of Public Health Medicine, Faculty of Medicine, Universiti Kebangsaan Malaysia, E-mail: drrose@ukm.edu.my.*

Received : January 26, 2025
Accepted : November 7, 2025
Published: November 28, 2025

The Maslach Burnout Inventory (MBI), a widely used tool for assessing burnout, evaluates the condition across three dimensions.[11] In addition, the MBI has been subject to considerable conceptual and psychometric criticism. These include concerns about weak factorial validity, ambiguous item phrasing, and significant overlap with symptoms of depression.[12] A further issue is its treatment of exhaustion, cynicism, and reduced efficacy as distinct domains rather than interrelated features of a single syndrome, which challenges its ability to capture burnout as a cohesive construct.[13] Moreover, the tool was originally designed for service-oriented occupations, limiting its applicability in healthcare-specific contexts.

The limitation highlights the critical need to develop a dedicated assessment tool to identify burnout among HCWs in their demanding environments effectively.[14] Questionnaires serve as a common and essential tool in these studies, enabling the collection of diverse information from participants.[15] Effective questionnaire design requires thorough planning to ensure that the questions and items accurately capture the specific constructs under investigation.[16] Following the design process, the development, design, and validation of a questionnaire require a rigorous and systematic approach to ensure reliable and meaningful responses and outcomes.[17]

Factor analysis is a statistical technique commonly employed to identify patterns within a set of variables by simplifying the relationships among them.[18] Its primary goal is to facilitate straightforward data interpretation. Initially adopted by educational researchers and psychologists to interpret self-report questionnaires, it has since expanded into a wide range of disciplines, including behavioral and social sciences, medical and nursing research, economics, and geography, particularly with advances in computing technology.[19] Since its development over a century ago, Exploratory Factor Analysis (EFA) has become a cornerstone in psychological research and one of the most widely applied statistical methods.[20] EFA is typically used in the early stages of scale development to identify underlying data patterns. It reduces numerous latent constructs into broader dimensions, helping align findings with theoretical frameworks by grouping related items.

Reliability is fundamental to ensuring the consistency of measurement validity.[21] The Rasch model, commonly referred to as the Item Response Theory or the Latent Trait Model, has gained recognition as a robust approach to evaluating the validity and reliability of research instruments. In recent years, this model has provided an alternative framework for understanding the measurement process and assessing the quality of instruments or questionnaires.[22] By applying the Rasch model, researchers can develop reliable and valid instruments. This model offers a powerful solution to address validity issues by providing useful statistical insights and valuable opportunities to evaluate the instrument quality.[23] Ensuring instrument validity and reliability is crucial for maintaining measurement accuracy and consistency.[24] The application of the Rasch model strengthens this process by ensuring that the instrument effectively captures the intended constructs, thus supporting the acquisition of valid and reliable results.

This study aimed to evaluate the validity and psychometric properties of a newly developed burnout questionnaire specifically designed for HCWs. The evaluation includes assessing the reliability of the collected data, analyzing the functioning of the response categories, and examining the instrument's dimensionality to ensure it accurately captures burnout in the target population. This study was expected to produce a psychometrically reliable and contextually appropriate instrument specifically designed for HCWs, effectively addressing significant measurement gaps and facilitating more accurate burnout identification, monitoring, and intervention.

Method

The development and validation of the Burnout Scale for Assessing Scale in Healthcare Workers (BSAS-HCW) followed a structured three-phase process in line with the instrument development guidelines proposed by Boateng[25]: (i) item development, (ii) scale development and validation, and (iii) scale evaluation. This study focused on the second phase, involving the instrument's psychometric validation using both EFA and the Rasch Measurement Model. Item generation was grounded in an extensive literature review of existing burnout measurement frameworks, particularly the MBI and Copenhagen Burnout Inventory. These instruments provided theoretical guidance for identifying the most relevant burnout domains for HCWs. Based on this foundation and supported by a prior systematic review and expert focus group discussions (FGDs), two core domains were selected: emotional exhaustion (EE) and cognitive exhaustion (CE). These dimensions were identified as particularly relevant in the Malaysian healthcare context, given local occupational stressors.

An initial pool of 20 items was created, with 10 items representing each domain, to reflect the theoretical constructs of burnout as experienced by the target population. Although the instrument was designed for HCWs, it was initially developed in English. The choice was informed by the high English proficiency of the target population, particularly medical doctors and nurses who routinely use English in clinical documentation, professional training, and workplace communication. English is also the principal medium of instruction in medical education throughout Malaysia.[26] To ensure that the instrument was linguistically accessible and culturally appropriate, cognitive debriefing sessions were conducted. Across seven sessions, item wording was refined based on participant feedback; for example, "overwhelmed" became "burdened," and "irritable" became "frustrated" to enhance clarity without altering meaning.

Before administering on a larger scale, a pilot test was conducted using cognitive debriefing interviews with a purposive sample of 12 HCWs, including house officers (medical interns) and staff nurses from clinical departments in Johor, Malaysia. These participants were recruited through department gatekeepers, and structured interviews were conducted to evaluate the clarity, comprehensibility, and emotional impact of the items. The interviews employed five cognitive probes, as recommended by Karlsson[27] to examine participants' interpretations of the items and identify any confusion or ambiguity. Based on the feedback obtained, several items were refined to enhance clarity while preserving their conceptual focus, thereby strengthening the validity of participant responses.

Following items refinement, content validity was assessed using the Fuzzy Delphi Method. A panel of eight experts in occupational health, psychiatry, psychology, and public health evaluated each item for importance, suitability, and relevance. The panel included participants from the earlier FGD, and additional professionals identified through academic networks. Experts rated each item using a 7-point Likert scale, and fuzzy scores were calculated to establish the level of consensus. Items were retained if they met three predetermined criteria: a threshold value (d) less than or equal to 0.2, expert agreement of 75% or more, and an alpha-cut value of at least 0.5. This method provided a transparent and systematic approach for refining the scale based on expert consensus.

Participants for the main validation study were recruited via convenience sampling, with support from clinical administrators at one of the government hospitals and the Johor Bahru District Health Office in Malaysia. Inclusion criteria ensured participants had relevant experience with occupational stressors. Eligible participants had to be currently employed HCWs with at least one year of professional experience and a weekly workload of 40 hours or more. A total of 245 responses were collected using an online survey administered through Google Forms, yielding a sample deemed sufficient for psychometric analysis and broadly representative of HCWs.

The BSAS-HCW instrument comprised 20 items evenly divided into two domains: EE and CE. Each item was rated on a 5-point Likert scale with response options including "Never" (1), "Rarely" (2), "Sometimes" (3), "Often" (4), and "Always" (5). This rating scale was selected for its simplicity and demonstrated reliability in psychological assessment tools. The psychometric properties of the instrument were first evaluated using an EFA in IBM SPSS Version 26 (SPSS-26-XYZ123456). Before extracting factors, the data were assessed using Bartlett's Test of Sphericity and the Kaiser-Meyer-Olkin (KMO) measure to determine sampling adequacy. Principal axis factoring was employed for extraction, followed by Promax rotation to account for the expected correlation between burnout constructs. Items with factor loadings below 0.40 were excluded, and factors with eigenvalues greater than 1.0 were retained.

The reliability of each factor was assessed using Cronbach's alpha, with values of 0.70 and above considered acceptable. To further evaluate the measurement properties of the scale, a Rasch analysis was conducted using Winsteps Version 3.72.3(WST-3723-ABC789101). Prior to analysis, the dataset was screened to remove misfitting respondents, defined as those with negative point-measure correlations, outfit mean-square values exceeding 2.0, or standardized residuals greater than 2. The Rasch model was used to assess item and person fit, rating scale performance, item-person targeting, person separation reliability, and differential item functioning (DIF). Acceptable model fit was indicated by person separation indices above 1.5 and person reliability values above 0.61. DIF analysis was conducted to identify items that may function differently across subgroups, such as sex or seniority level. Items were flagged if they demonstrated a logit difference greater than 0.50 and a t-value of 2 or more. To mitigate potential bias throughout the instrument development process, several precautions were taken. Expert panels were composed of individuals from diverse clinical and academic backgrounds to minimize subjectivity. DIF analysis further helped identify and correct any items that might have unfairly favored specific demographic subgroups. These multi-layered strategies contributed to the development of a valid, reliable, and contextually appropriate tool for assessing burnout among HCWs.

Results

The psychometric properties of the BSAS-HCW were evaluated using both EFA and Rasch measurement modeling via Winsteps software. A total of 245 HCWs completed the 20-item questionnaire, as detailed in Tables 1 and 2. The majority were female (204, 83.3%) and of Malay ethnicity (170, 69.4%). Participants came from a wide range of departments, with the largest groups working in paediatrics (60, 24.5%) and medical wards (58, 23.7%). In terms of occupation, staff nurses comprised the largest subgroup (141, 57.7%), followed by house officers (35, 14.3%) and medical officers (33, 13.5%). Data screening identified three misfitting respondents based on established Rasch criteria (Table 3): Outfit Mean Square (MNSQ) values above 2.0, Z-standardized values (ZSTD) exceeding ±2, and negative point-measure correlations. Specifically, Entry 33 (Outfit MNSQ = 4.43; ZSTD = 6.46; Correlation = –0.01), Entry 2 (Outfit MNSQ = 3.43; ZSTD = 4.56; Correlation = –0.14), and Entry 22 (Outfit MNSQ = 2.76; ZSTD = 4.20; Correlation = –0.07) were excluded from further analysis to improve model fit and reliability.

Following data cleaning, Rasch analysis confirmed the instrument's structural soundness and measurement precision (Table 4). Dimensionality assessment through principal component analysis of standardized residuals revealed that the EE scale explained 75.4% of the total variance, and the CE scale explained 69.4%, both values well above the 40% benchmark for unidimensionality. The first residual contrast was 1.83 eigenvalue units for EE and 2.22 for CE, the latter suggesting minor multidimensionality but still within acceptable limits when assessed against item variance ratios. Point-measure correlations ranged from 0.86 to 0.89 (EE) and 0.71 to 0.87 (CE), and all item fit statistics (infit and outfit MNSQ) were within the acceptable range of 0.6–1.4, confirming strong alignment with the Rasch model.

The scale's 5-point Likert-type rating categories functioned effectively, as confirmed by Winsteps' rating scale diagnostics. Each category had more than 10% endorsements, and observed averages increased monotonically. Andrich thresholds progressed in an orderly sequence across categories, and probability curves showed distinct peaks, which demonstrated that each category captured a unique level of the underlying trait. Reliability indices were excellent, with person reliability of 0.95 (EE) and 0.93 (CE), and item reliability of 0.95 (EE) and 0.98 (CE). Separation indices also exceeded minimum requirements, with person separation indices of 4.37 (EE) and 3.67 (CE), confirming the scale's capacity to distinguish among varying levels of burnout. DIF analysis, conducted in Winsteps, detected only minor group-related differences. For EE, items EE3 and EE6 showed small DIF effects between sexes. For CE, items CE3, CE5, and CE10 displayed slight variations across sexes and racial subgroups. However, none of the items exceeded the logit difference threshold of ±0.50 or t-value threshold of ±2, indicating minimal measurement bias and supporting cross-group validity.

Table 1. Participants' Sociodemographic (N= 245)

Variable	Category	n (%)
Race	Malay	170 (69.4)
	Chinese	34 (13.9)
	Indian	22 (9.0)
	Others	19 (7.8)
Department	Medical	58 (23.7)
	Paediatric	60 (24.5)
	Emergency & Traumatology	14 (5.7)
	Surgery	24 (9.8)
	Obstetrics & Gynaecology	32 (13.1)
	Anaesthesiology	8 (3.3)
	Orthopaedic	13 (5.3)
	Cardiology	7 (2.9)
	Neurosurgery	6 (2.4)
	Pharmacy	10 (4.1)
	Polycystic Kidney Disease	5 (2.0)
	Dermatology	1 (0.4)
	Urology	6 (2.4)
	Ophthalmology	1 (0.4)
Occupation	Medical Officer	33 (13.5)
	House Officer	35 (14.3)
	Specialist/Consultant	14 (5.7)
	Pharmacist	11 (4.5)
	Medical Assistant	11 (4.5)
	Staff Nurse	141 (57.7)
Sex	Male	41 (16.7)
	Female	204 (83.3)

Table 2. Cognitive and Emotional Exhaustion Items of the BSAS-HCW

	Cognitive Exhaustion		Emotional Exhaustion
CE 1	How often do you feel mentally drained at the end of your workday?	EE1	How often do you feel emotionally drained by your work?
CE 2	How frequently do you find it difficult to concentrate on your tasks?	EE2	How frequently do you feel overwhelmed by your emotional responses to work situations?
CE 3	How often do you have trouble remembering important details at work?	EE3	How often do you feel less enthusiastic about your work?
CE 4	How often do you feel overwhelmed by the difficulty of your work tasks?	EE4	How frequently do you experience feelings of hopelessness regarding your work?
CE 5	How often do you make mistakes in your work?	EE5	How often do you feel frustrated with the demands of your work?
CE 6	How frequently do you feel unable to think clearly during your shift?	EE6	How frequently do you feel sad or down because of your work?
CE 7	How often do you experience "brain fog" or confusion while working?	EE7	How often do you feel detached or disconnected from your work?
CE 8	How frequently do you have trouble concentrating on one task without being distracted?	EE8	How frequently do you feel irritable or angry due to occupational stress?
CE 9	How often do you feel your decision-making abilities are impaired due to mental exhaustion?	EE9	How often do you feel you have no energy to engage with colleagues or patients?
CE 10	How frequently do you find it hard to stay mentally alert throughout your entire shift?	EE10	How frequently do you find it difficult to recover emotionally from a tough workday?

Table 3. Analysis of Misfit Persons

Entry Number	Total Score	Total Count	JMLE Measure	Model S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PTMEASUR-AL	Correlation	Observed (%)	Expected (%)
33	36	20	-2.54	0.38	4.28	6.18	4.43	6.46	A-.01	0.31	10	59.6
2	46	20	-1.23	0.35	3.44	4.75	3.43	4.56	C-.14	0.33	30	57.4
22	35	20	-2.69	0.38	2.76	4.14	2.76	4.2	G-.07	0.3	20	58.2

Notes: JMLE Measure = Joint Maximum Likelihood Estimation Measure, Model S.E. = Model Standard Error, Infit MNSQ = Information-weighted Mean Square, Infit ZSTD = Information-weighted Standardized Z-score, Outfit MNSQ = Outlier-sensitive Mean Square, Outfit ZSTD = Outlier-sensitive Standardized Z-score; PTMEASUR-AL Correlation = Point-Measure Correlation.

Table 4. Rasch Model Analysis: Construct Validation

Evaluation Domain	Criteria	Emotional Exhaustion	Cognitive Exhaustion
Dimensionality Assessment	≥40% variance explained (PCA)	✓ 75.4%	✓ 69.4%
	First contrast <2 eigenvalues	✓ 1.83	⚠ 2.22 (Slight multidimensionality)
	Item variance/first contrast ≤3	✓ Yes	✓ Acceptable
Item Fit Metrics	PT-measure correlation 0.4–0.8	✓ (0.86–0.89)	✓ (0.71–0.87)
	Infit/Outfit MNSQ 0.6–1.4	✓ (0.86–1.23)	✓ (0.62–1.38)
	ZSTD range −2 to +2	✓ All within range	✓ Minor deviations acceptable
Rating Scale Validation	≥10 responses per category	✓ Yes	✓ Yes
	Monotonic threshold ordering (≥1 logit advance)	✓ Ordered	✓ Ordered
	Probability curves show distinct peaks	✓ Clear peaks	✓ Clear peaks
	Outfit MNSQ for categories <2	✓ Yes	✓ Yes
Differential Item Functioning	Logit diff >0.5 & t ≥2 → flag item	⚠ Minor DIF (EE3, EE6)	⚠ Minor DIF (CE3, CE5, CE10)
Reliability & Targeting Metrics	Person Separation ≥1.5	✓ 4.37	✓ 3.67
	Item Reliability	✓ 0.95	✓ 0.98
	Person Reliability	✓ 0.95	✓ 0.93
	Wright Map: Items well targeted	✓ Balanced distribution	✓ Slight item skew

Notes: PCA = principal component analysis, MNSQ = Mean Square, ZSTD = z-standardized value, DIF = differential item functioning, CE = cognitive exhaustion, EE = emotional exhaustion.

The KMO measure of sampling adequacy was 0.973, and Bartlett's Test of Sphericity was statistically significant (χ^2 = 5434.216, df = 190, p-value <0.01), confirming the suitability of the data for factor analysis (Table 5). Principal axis factoring with a Promax rotation revealed a clear two-factor structure, accounting for 74.80% of the total variance. Factor 1 initially accounted for 69.85%, and Factor 2 for 4.95%. After rotation, Factor 1 explained 45.37% and Factor 2 explained 29.43%. All items demonstrated positive factor loadings, ranging from 0.416 to 0.816, indicating direct and meaningful relationships with their respective constructs (Table 6). Descriptive statistics further supported the validity of the scale. Mean scores for the CE items ranged from 2.46 to 3.47, with standard deviations between 0.964 and 1.287. For EE, the item means ranged from 2.71 to 3.25. Item-total correlations exceeded 0.61 for all items, and internal consistency was excellent, with Cronbach's alpha values of 0.949 (CE) and 0.971 (EE). These findings indicate that participants reported moderate levels of exhaustion, with EE being slightly more prevalent than CE among the sampled HCWs.

Table 5. Kaiser-Meyer-Olkin and Bartlett's Test

Statistic	Value
Kaiser-Meyer-Olkin Measure of Sampling Adequacy	0.973
Bartlett's Test of Sphericity (Chi-Square)	5434.216
Degrees of Freedom (df)	190
Significance (Sig.)	0.01

Table 6. Factor Loadings from Principal Axis Factor Analysis

Factor	Item	Mean (SD)	Loadings	Item-Total Correlation	Cronbach's Alpha
Cognitive Exhaustion	CE1	3.47 (1.183)	0.416	0.788	0.949
	CE2	2.87 (1.080)	0.629	0.802	
	CE3	2.86 (1.082)	0.738	0.668	
	CE4	3.22 (1.178)	0.476	0.773	
	CE5	2.46 (0.964)	0.743	0.614	
	CE6	2.68 (1.145)	0.782	0.834	
	CE7	2.70 (1.188)	0.731	0.805	
	CE8	2.61 (1.161)	0.724	0.777	
	CE9	2.92 (1.246)	0.614	0.843	
	CE10	2.86 (1.287)	0.604	0.867	
Emotional Exhaustion	EE1	3.25 (1.247)	0.811	0.865	0.971
	EE2	3.07 (1.195)	0.816	0.848	
	EE3	2.92 (1.241)	0.754	0.842	
	EE4	2.76 (1.255)	0.793	0.868	
	EE5	3.15 (1.270)	0.810	0.871	
	EE6	2.98 (1.253)	0.816	0.856	
	EE7	2.71 (1.261)	0.765	0.819	
	EE8	3.11 (1.224)	0.812	0.846	
	EE9	2.92 (1.259)	0.745	0.849	
	EE10	2.87 (1.263)	0.794	0.854	

Notes: EE = emotional exhaustion ; CE = cognitive exhaustion, SD = standard deviation

To evaluate scale targeting and item difficulty, Wright Maps were generated using Winsteps for the EE and CE scales. These maps plot respondent ability and item difficulty on a shared logit scale. For the EE scale, item difficulty was generally well-aligned with respondent distribution, though slight gaps at the extremes suggest reduced precision for very low or high burnout levels. On the CE scale, respondents were more tightly clustered around the mean, with CE5 identified as the most difficult item and CE1 as the easiest. Overall, the maps indicate good item-person targeting and confirm that the BSAS-HCW captures a broad range of burnout levels among healthcare workers.

Discussion

This study assessed two primary constructs emerging from the item development phase, which incorporated both inductive and deductive approaches. The inductive method focused on identifying themes and patterns derived from qualitative insights. At the same time, the deductive approach relied on existing theoretical frameworks, literature reviews, and established constructions to inform and guide the item creation process. CE does not imply a total inability to solve problems; on the contrary, deficiencies tend to manifest primarily in new and challenging situations that require non-standard, flexible coping strategies. Furthermore, CE is often accompanied by motivational deficits, leading to cognitive demobilization and reduced motivation to exert mental effort, particularly following helplessness training.[28] A previous study suggested that CE develops through sustained exposure to uncontrollable situations, a process exemplified by helplessness-training experiments involving repeated exposure to unsolvable tasks.[29] Prolonged CE may contribute to reactive depression by disrupting information processing and reinforcing negative thought patterns, such as consistently attributing negative events to personal failure or unchangeable causes.[30]

When individuals are repeatedly exposed to emotionally demanding situations, they may experience mental and physical exhaustion.[31] A key outcome of prolonged stress was EE, occurring when a person's internal emotional resources are no longer sufficient to cope with work demands. This condition is often considered the first and most central stage of burnout. EE has attracted increasing attention from organizational researchers due to its negative impact on job satisfaction, quality of work life, and overall organizational performance. It is particularly likely to occur in environments with excessive task demands and inadequate support or resources, ultimately undermining employee well-being and productivity.[31]

Rasch modelling is not intended to replace traditional psychometric methods for evaluating construct validity, but to complement them.[32] It offers distinct advantages by allowing researchers to assess whether the items in a scale collectively measure a single underlying construct (unidimensionality). Additionally, Rasch analysis can detect statistically unusual (idiosyncratic) response patterns, calibrate the relative difficulty or intensity of items on an interval scale, and provide individual-level measurements along this newly constructed variable.[32] Rasch modelling generates both person and item separation indices, offering measures for individual items and respondents. The person measure reflects each individual's perceived position regarding the underlying trait being assessed. Additionally, person-fit analysis assesses whether an individual's responses align consistently with the expectations of the Rasch model. The primary aim of EFA is to simplify data, making it easier to interpret and identify relationships and patterns among the observed variables within a measurement tool.[33] In essence, EFA reorganizes observed variables into smaller groups, represented by a reduced number of latent variables that are not directly observable but have the same variance.

The EFA results confirmed the presence of two distinct but related dimensions. Items from the CE scale exhibited variable performance. CE6 ("How frequently do you feel unable to think clearly during your shift?") achieved the highest factor loading (0.782), strongly representing CE. Similarly, CE7 ("How often do you experience brain fog or confusion?") demonstrated robust validity with a loading of 0.731. CE5 ("How often do you make mistakes in your work?"), despite showing a strong factor loading (0.743), showed a weaker item-total correlation (0.614), suggesting lower internal consistency. This difference may reflect the medicolegal and cultural sensitivities surrounding error reporting in Malaysia, where admitting errors can result in criminal and professional consequences. A study in Taiwan has shown that HCWs are less likely to report errors due to punitive consequences or fear of reputational damage.[34] Moreover, the increasing global emphasis on patient safety protocols may discourage honest responses, reducing the reliability of error-related items.[35] Despite these challenges, CE5 retained its conceptual value and was retained in the scale, albeit with recognition of its potential response bias.

The EE scale performed more consistently, with all items demonstrating strong factor loadings. EE3 ("How often do you feel less enthusiastic about your work?"), EE4 ("How frequently do you experience feelings of hopelessness regarding your work?"), and EE5 ("How often do you feel frustrated with the demands of your work?") reflected high alignment with the construct, consistent with the well-documented emotional strain reported among HCWs. These findings paralleled international studies. For example, qualitative studies in Malaysia and Bangladesh have highlighted reduced enthusiasm, hopelessness, and frustration as common emotional consequences of healthcare work, particularly during the COVID-19 pandemic.[36,37] Similar patterns are reported in the Western contexts, where EE has long been recognized as a central component of burnout in MBI.[38] These results suggested that while the CE scale requires contextual sensitivity, the EE scale demonstrated robust cross-cultural validity.

Rasch analysis, complementing the EFA, confirmed the tool's validity and reliability. The CE and EE scales explained 69.4% and 75.4% of the variance, respectively, both surpassing the 40% benchmark for unidimensionality. Item and person separation indices exceeded 3.5 for both scales, indicating strong differentiation across exhaustion levels. Minor misfits in CE3 and CE5 were within acceptable limits and did not affect overall model fit. The DIF analysis showed minimal differences by sex and race, aligning with international findings that report only minor subgroup differences in burnout measurement.[39] The rating scale also functioned effectively, with ordered thresholds and distinct category probability curves, reinforcing the scale's measurement robustness. This study offered important practical implications. The BSAS-HCW presented a psychometrically sound alternative to the MBI, particularly in contexts where cultural and medicolegal factors may influence responses. While the MBI is widely used, its focus on depersonalization and personal accomplishment has been criticized for limited cultural relevance in Asian healthcare settings. However, this study had limitations. Data were drawn from only two Malaysian institutions, limiting generalizability. Additionally, its cross-sectional design precluded assessment of temporal changes. Future research should adopt longitudinal methods and conduct cross-cultural validation to enhance the scale's applicability and robustness across diverse healthcare contexts.

Conclusion

This study highlights the successful development and validation of a burnout assessment tool specifically designed for HCWs, addressing the unique challenges of their stressful environments. Using EFA and the Rasch model, the instrument demonstrates strong psychometric properties, including reliability, unidimensionality, and internal consistency across its two dimensions: cognitive and emotional exhaustion. These findings emphasize the broad impact of burnout on HCWs' emotional resilience and cognitive functioning and underscore the importance of a rigorous statistical approach to refining measurement tools. This validated instrument provides a reliable basis for identifying burnout, enabling more effective interventions to support HCWs' well-being.

Abbreviations

HCWs: Healthcare Workers; MBI: Maslach Burnout Inventory; EFA: Exploratory Factor Analysis; BSAS-HWC: Burnout Scale for Assessing Scale in Healthcare Workers; FGD: focus group discussion; EE: emotional exhaustion; CE: cognitive exhaustion; KMO: Kaiser-Meyer-Olkin; DIF: differential item functioning; MNSQ: Mean Square; ZTSD: Z-standardized values; SD: Standard Deviation.

Ethics Approval and Consent to Participate

This study received ethical approval from the Ethics Committee of the National University of Malaysia (JEP-2023-843) and the Malaysian Ministry of Health. Informed consent was obtained from all participants after providing a brief explanation of the study's objectives.

Competing Interest

The authors declare no conflict of interest. This study was conducted independently with no commercial or financial relationships that could be construed as a potential conflict of interest.

Availability of Data and Materials

The data supporting the findings of this study are available upon reasonable request from the corresponding author. Additional information, including the Supplementary File, provides accessible details on the datasets used in this research.

Authors' Contribution

RI contributed to the study design and data analysis, and SKV was involved in data collection, statistical analysis, and drafting the methodology section and manuscript preparation. BMS provided critical revisions, contributed to the theoretical framework, and reviewed the final manuscript.

Acknowledgment

The authors gratefully acknowledge the support of the Ministry of Health Malaysia and the Ethics Committee of the National University of Malaysia for their approval and guidance, as well as the participating healthcare workers for their invaluable contribution to this study.

References

1. Mehta V, Ajmera P, Kalra S, et al. Human resource shortage in India's health sector: A scoping review of the current landscape. BMC Public Health. 2024; 24: 1368. DOI: 10.1186/S12889-024-18850-X.
2. Vallasamy SK, Ismail R, Sundram BM. Effectiveness of Burnout Assessment Scales Among Healthcare Workers in Malaysia: A Systematic Review. J Health Res. 2024; 38 (4): 350–369. DOI: 10.56808/2586-940X.1089.
3. Jachmann A, Loser A, Mettler A, et al. Burnout, Depression, and Stress in Emergency Department Nurses and Physicians and the Impact on Private and Work Life: A Systematic Review. JACEP Open. 2025; 6 (2): 100046. DOI: 10.1016/J.ACEPJO.2025.100046.
4. Kabunga A, Kigongo E, Musinguzi M, et al. Level of burnout and associated factors among healthcare workers in central Uganda: A facility-based cross-sectional study. PLoS One. 2024; 19 (10): e0309701. DOI: 10.1371/JOURNAL.PONE.0309701.
5. Woo T, Ho R, Tang A, et al. Global prevalence of burnout symptoms among nurses: A systematic review and meta-analysis. J Psychiatr Res. 2020; 123: 9–20. DOI: 10.1016/J.JPSYCHIRES.2019.12.015.
6. Shen X, Xu H, Feng J, et al. The global prevalence of burnout among general practitioners: A systematic review and meta-analysis. Fam Pract. 2022; 39 (5): 943–950. DOI: 10.1093/FAMPRA/CMAB180.
7. Ong J, Ong AML, Ong S, et al. Burnout and work-related stressors in gastroenterology: A protocol for a multinational observational study in the ASEAN region. BMJ Open Gastroenterol. 2020; 7: ee000534. DOI: 10.1136/BMJGAST-2020-000534.
8. Abdul Aziz AF, Ong T. Prevalence and associated factors of burnout among working adults in Southeast Asia: Results from a public health assessment. Front Public Health. 2024; 12: 1326227. DOI: 10.3389/FPUBH.2024.1326227.
9. Chor WPD, Ng WM, Cheng L, et al. Burnout amongst emergency healthcare workers during the COVID-19 pandemic: A multi-center study. Am J Emerg Med. 2021; 46: 700–702. DOI: 10.1016/j.ajem.2020.10.040.
10. Hart D, Paetow G, Zarzar R. Does implementation of a corporate wellness initiative improve burnout? West J Emerg Med. 2019; 20: 138–144. DOI: 10.5811/WESTJEM.2018.10.39677.
11. Maslach C, Leiter MP. Understanding the burnout experience: Recent research and its implications for psychiatry. World Psychiatry. 2016; 15 (2): 103–111. DOI: 10.1002/WPS.20311.
12. Bianchi R, Swingler G, Schonfeld IS. The Maslach Burnout Inventory is not a measure of burnout. Work. 2024; 79 (3): 1525–1527. DOI: 10.3233/WOR-240095.
13. Bianchi R, Schonfeld IS, Laurent E. Burnout-depression overlap: A review. Clin Psychol Rev 2015; 36: 28–41. DOI: 10.1016/J.CPR.2015.01.004.
14. Hochberg CH, Eakin MN. Keys to Successful Survey Research in Health Professions Education. ATS Sch. 2024; 5 (1): 206–217. DOI: 10.34197/ATS-SCHOLAR.2023-0112RE
15. Brant JM, Haas-Haseman ML, Wei SH, et al. Understanding and Evaluating Survey Research. J Adv Pract Oncol. 2015; 6 (2): 168–171. DOI: 10.6004/jadpro.2015.6.2.9.
16. Yusoff MSB, Arifin WN, Hadie SNH. ABC of questionnaire development and validation for survey research. Educ Med J. 2021; 13 (1): 97–108; DOI: 10.21315/EIMJ2021.13.1.10.
17. Koo M, Yang S-W. Questionnaire Use and Development in Health Research. Encyclopedia. 2025; 5 (2): 65. DOI: 10.3390/ENCYCLOPEDIA5020065.
18. Tavakol M, Wetzel A. Factor Analysis: A means for theory and instrument development in support of construct validity. Int J Med Educ. 2020; 11: 245–247. DOI: 10.5116/IJME.5F96.0F4A.
19. Shrestha N. Factor Analysis as a Tool for Survey Analysis. Am J Appl Math Stat. 2021; 9 (1): 4–11. DOI: 10.12691/AJAMS-9-1-2.
20. Vakili MM. Assessment of construct validity questionnaires in psychological, educational and Health research: Applications, Methods, and Interpretation of Exploratory factor analysis. J Med Educ Dev. 2018; 11 (30): 4–21. DOI: 10.29252/EDCJ.11.30.4.

21. Yasin RM, Yunus FAN, Rus RC, et al. Validity and Reliability Learning Transfer Item Using Rasch Measurement Model. Procedia Soc Behav Sci. 2015; 204: 212–217. DOI: 10.1016/J.SBSPRO.2015.08.143.
22. Lalot F, Räikkönen J, Ahvenharju S. An item response theory approach to measurement in environmental psychology: A practical example with environmental risk perception. J Environ Psychol. 2025; 101: 102520. DOI: 10.1016/J.JENVP.2025.102520.
23. Davis DR, Boone W. Using Rasch analysis to evaluate the psychometric functioning of the other-directed, lighthearted, intellectual, and whimsical (OLIW) adult playfulness scale. Int J Educ Res Open. 2021; 2: 100054. DOI: 10.1016/J.IJEDRO.2021.100054.
24. Avinç E, Doğan F. Digital literacy scale: Validity and reliability study with the rasch model. Educ Inf Technol (Dordr). 2024; 29 (17): 22895–22941. DOI: 10.1007/S10639-024-12662-7.
25. Boateng GO, Neilands TB, Frongillo EA, et al. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. Front Public Health. 2018; 6: 366616. DOI: 10.3389/FPUBH.2018.00149.
26. Teng ZS, Ser GTZ, Hong WH, et al. Malaysian Medical Students' Career Intention (MMSCI): A cross-sectional study. Hum Resour Health. 2024; 22: 59. DOI: 10.1186/S12960-024-00939-4.
27. Karlsson N, Atkinson MJ, Müllerová H, et al. Validation of a diagnosis-agnostic symptom questionnaire for asthma and/or COPD. ERJ Open Res. 2021; 7 (1): 00828-2020. DOI: 10.1183/23120541.00828-2020.
28. Józefczyk A. The Effects of Stability of Uncontrollability Experiences on Cognitive Exhaustion. Soc Psychol Bull. 2018; 13 (1): e2421. DOI: 10.5964/SPB.V13I1.25498.
29. Tafet GE, Ortiz Alonso T. Learned helplessness and learned controllability: From neurobiology to cognitive, emotional and behavioral neurosciences. Front Psychiatry. 2025; 16: 1600165. DOI: 10.3389/FPSYT.2025.1600165.
30. Machado S, Coelho DB, Englert C, et al. Shared Demands Between Cognitive and Physical Tasks May Drive Negative Effects of Fatigue: A Focused Review. Front Sports Act Living. 2019; 1: 45. DOI: 10.3389/FSPOR.2019.00045.
31. Jin MJ, Kim JS, Lee HS, et al. The Impact of Emotional Exhaustion on Psychological Factors in Workers with Secondary Traumatic Experiences: A Multi-Group Path Analysis. Psychiatry Investig. 2020; 17 (11): 1064–1072. DOI: 10.30773/PI.2019.0313.
32. Boone WJ. Rasch Analysis for Instrument Development: Why, When, and How? CBE Life Sci Educ. 2016; 15 (4): rm4. DOI: 10.1187/CBE.16-04-0148.
33. Sürücü L, YIKILMAZ İ, MASLAKÇI A. Exploratory Factor Analysis (EFA) in Quantitative Researches and Practical Considerations. OSF Preprints. 2024; 13 (2): 947–965. DOI: 10.31219/OSF.IO/FGD4E.
34. Tsai SF, Wu CL, Ho YY, et al. Medical malpractice in hospitals-how healthcare staff feel. Front Public Health. 2023; 11: 1080525. DOI: 10.3389/FPUBH.2023.1080525.
35. AlBalushi AA, Al-Asmi A, Al-Shekaili W, et al. Medical malpractice in Oman: A 12-year retrospective record review. PLoS One. 2023; 18 (8): e0290349. DOI: 10.1371/JOURNAL.PONE.0290349.
36. Roslan NS, Yusoff MSB, Asrenee AR, et al. Burnout Prevalence and Its Associated Factors among Malaysian Healthcare Workers during COVID-19 Pandemic: An Embedded Mixed-Method Study. Healthcare. 2021; 9 (1): 90. DOI: 10.3390/HEALTHCARE9010090.
37. Razu SR, Yasmin T, Arif TB, et al. Challenges Faced by Healthcare Professionals During the COVID-19 Pandemic: A Qualitative Inquiry From Bangladesh. Front Public Health. 2021; 9: 647315. DOI: 10.3389/FPUBH.2021.647315.
38. Maslach C, Jackson SE. The Measurement of Experienced Burnout Maslach Burnout. J Org Behav. 1981; 2 (2): 99–113. DOI: 10.1002/job.4030020205.
39. Huang TW, Wu PC, Mok MMC. Examination of Gender-Related Differential Item Functioning Through Poly-BW Indices. Front Psychol. 2022; 13: 821459. DOI: 10.3389/FPSYG.2022.821459.