

## Research Article

# The Development and Psychometric Validation of Culture of Patient Safety Scale Under Rasch Objective Measurement Theory

Odunayo Kolawole Omolade (He/Him) <sup>1</sup> and John Stephenson<sup>2</sup>

<sup>1</sup>School of Health, Education, Policing and Science, University of Staffordshire, Blackheath Lane, Stafford, UK

<sup>2</sup>School of Human & Health Sciences, University of Huddersfield, Queensgate, Huddersfield, UK

Correspondence should be addressed to Odunayo Kolawole Omolade; [odunayokolawole.omolade@staffs.ac.uk](mailto:odunayokolawole.omolade@staffs.ac.uk)

Received 3 December 2024; Accepted 25 February 2025

Academic Editor: Majed Alamri

Copyright © 2025 Odunayo Kolawole Omolade and John Stephenson. International Journal of Clinical Practice published by John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

**Background:** Assessing the culture of patient safety in healthcare settings is pivotal for continuously reinforcing effective, safe and quality patient care. However, most of the rating scales lack evidence of objective validation of the measuring instruments.

**Aim:** To determine the psychometric properties of culture of patient safety scale under the Rasch objective measurement theory.

**Method:** The validation of the culture of safety scale was underpinned by the four stages of rating scale development in Rasch objective measurement theory. The first stage involved literature review to shortlist items considered theoretically relevant to culture of patient safety in hospital settings. In the second stage, a panel of academic and practitioners individually reviewed the selected items to give external face validity based on professional experiences. In the third stage, 967 participants from public maternity settings in Nigeria voluntarily accessed the nine items forming the culture of patient safety scale online over 8-week period. Ethical approval was given by the nurses' association and University of Huddersfield. Subsequently, all the data were exported to SPSS and Winsteps Version 5.0.0.0 for evaluation of the psychometric assumptions. Essential psychometric properties evaluated are dimensionality, category functioning, item difficulty/agreeability, local independence, reliability and item validity. In the fourth stage, problematic items were identified and moderated based on the outcome of the measurement assumptions. Consequently, final decisions made included retention, modification or expulsion of items, making no meaningful contribution to the variable measurement.

**Conclusion and Implication:** The culture of safety scale has excellent psychometric properties and therefore recommended for use among practitioners and researchers. No direct contribution from the public or patients required in this study.

**Keywords:** culture; measurement; patient safety; psychometric; questionnaire; rating scale; validation assessment

## 1. Background

This study intends to apply the four stages of objective measurement theory (theoretical relevance, face validation, objective test of assumptions and item revision) [1] in developing and validating the culture of patient safety scale. In hospital settings, the culture of safety exerts a profound influence on quality of treatment by fostering an environment conducive to patient well-being and staff efficacy [2, 3]. In an article titled 'culture eats strategy every time', Melnyk emphasizes on the culture of best practices as the holy grail of patient safety in healthcare organizations [4]. A strong

culture of safety in maternity settings is crucial to significantly reduce the risk of preventable harm to mothers and infants, enhancing overall patient safety [5]. Secondly, safety culture enhances better communication and teamwork among healthcare providers, which is essential for managing emergencies and complex deliveries [5, 6]. Lastly, a positive safety culture improves staff morale and job satisfaction, leading to more consistent and high-quality care [6, 7]. These factors collectively contribute to better health outcomes by sustaining a supportive environment for both patients and healthcare professionals [8]. Consequently, questionnaires and surveys featuring evidence-based practices in making

treatment decision, hospital leadership, communication, teamwork and quality improvement are considered integral indicators of rating of patient safety [3, 9, 10]. Additionally, measuring culture of safety in maternity settings and elsewhere has gained traction in deepening the understanding and prediction of dynamic relationships among the human, ergonomic and institutional factors feeding into the concept of patient safety [11–13]. Frequently, avoidable harms caused to patients from preventable medical negligence or malpractices are associated with broken culture of safety in healthcare settings [13, 14]. However, the indicators of culture of patient safety in hospital settings are complex due to the multifaceted viewpoints of culture and safe practices in clinical areas [3, 13–15].

Despite increasing interest to gauge patient safety on a scale, measuring the culture of patient safety presents several challenges starting with the ambiguous definitions of patient safety [15]. Safety issues in clinical settings encompass various dimensions such as leadership, psychosocial environment and teamwork, thereby making a universal definition an almost impossible mission [11, 15]. This lack of standardization allows for other inconsistencies in the methodologies engaged in developing rating scales that measure culture of safety. For instance, the proliferations of quantitative surveys purportedly assessing culture of maternity safety have been criticized as an indication of confusion emerging from limited understanding of the concept of safety [7]. Moreover, already developed surveys under the classical methods are not evaluated for response bias due to failure to engage with techniques of objectively assessing the psychometric properties of the rating scale [9]. Even though, proponents of surveys of patient safety are aware that staff may provide socially desirable responses rather than honest feedback especially in environments where there is fear of retribution [9, 15]. Interestingly, little is known about efforts to address these challenges using advanced statistical techniques rooted in Rasch model of developing surveys that inculcate objective measurement theory. A major drawback of applying Rasch model in developing clinical survey or questionnaire is the complexity of the mathematical equation informing the model and associated reluctance among researchers [16, 17].

The result of a systematic review suggests that the common interest in using surveys to assess the culture of safety in clinical areas has not yielded any meaningful agreement in the methodologies to develop and validate the scales [7]. This poses significant difficulty for clinicians and stakeholders when selecting a rating scale since no agreed method is known to underpin the determination of the psychometric properties of the scales presented. After assessing 24 rating scales of culture of patient safety developed in the United Kingdom, Australia, Canada, Japan and Europe, a systematic review reported the classical test theory (CTT) methods for developing Likert-type surveys as the common attributes of all the scales [15]. Additional conventional (CTT) psychometric assessment identified include factor analyses, construct validity and the Cronbach's alpha values—projected as a proof of internal reliability of most of the scales [9]. Consequently, rigorous

assessment of psychometric properties of all the patient safety questionnaires or surveys used as rating scales is recommended [9, 15], thus indicating the interest to inculcate objective assumption test such as Rasch theory.

Addressing the challenges in developing rating scales is crucial so that skew results from poorly developed tools do not obscure true safety culture issues. Correspondingly, psychometric assessment under Rasch method is conducted so that the outcome of the assumption test may identify problematic items and guide the survey designer in making evidence-based decisions aimed at improving the overall function of the scale [18–20]. In summary, the core advantage of Rasch techniques over CTT methods is that the former objectively assess item quality and invariance of item measures using various diagnostic parameters, whereas the latter offer no mathematically sound solution to the problems [21]. Furthermore, placing both item difficulty and respondents' measures on a common (interval) logit unit scale allows the developer to evaluate the gaps and item redundancies on the rating scale [21–24]. On this background, the development of the culture of patient safety scale was informed by the four stages of Rasch technique of validating a rating scale as outlined in the method section below.

## 2. Material and Methods

We adopted the Rasch techniques of questionnaire validation simplified into four stages (see Figure 1 above) known as best practices in survey design [1, 16]. The first step emphasized on reviewing relevant theories to extract a set of items or identify existing questionnaire that addresses the same problem [1]. Correspondingly, we built on a relevant Likert-type questionnaire [25] by modifying the items to mirror the culture of patient safety in hospital settings. In the original study, the nine items ranked on three ordinal categories possess Cronbach's alpha score of 0.85 [25]. The major modifications to the scale include paraphrasing the items to mirror patient safety issues in hospitals and increasing the polytomous category measures from three to four (Strongly agree = 4, Agree = 3, Disagree = 2, Strongly disagree = 1) in Table 1.

*2.1. Step 2: Face Validity Through Panel Review.* Face validity, assessed by expert panel, ensures that a survey appears to measure what it claims to, enhancing credibility and acceptance among respondents, which is crucial for objective measurement theory [18, 26]. Expert knowledge and personal experience guide the nuanced calibration of the scale's content, shaping it into a tool that accurately measures, differentiates, and categorizes based on a set of items or indicators.

*2.1.1. Review Panel Constitution.* The study areas being maternal and child hospital settings in Nigeria, the members of the panel were nine including a professor of obstetrics, a consultant obstetrician, two senior lecturers in nursing and midwifery, the head of nursing services, a senior nursing

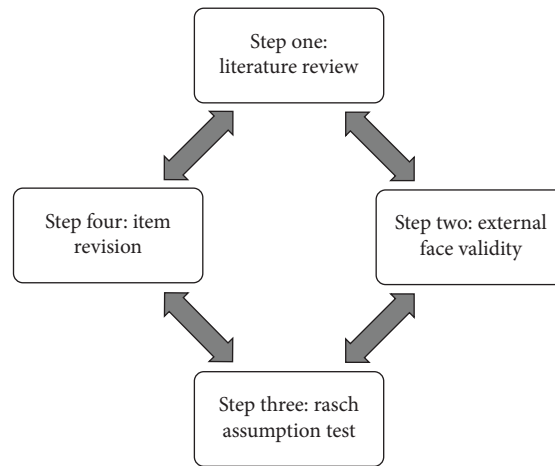


FIGURE 1: Four stages of survey design.

TABLE 1: Culture of patient safety scale.

Item	Strongly agree	Agree	Disagree	Strongly disagree
1. There is a good deal of teamwork where I work to implement the best practice protocol				
2. I am satisfied with the staff interactions in my hospital to promote best practice protocol				
3. Physicians, in general, cooperate with the nursing staff to adhere with treatment guidelines				
4. The patient safety team are dedicated to best practices within my hospital				
5. We conduct a routine review of treatment outcomes in my hospital				
6. We all have an equal voice in the multidisciplinary team in planning the treatment policies and procedures in my hospitals				
7. I am satisfied with the medical equipment used in patient treatment in my hospital				
8. My superior takes adherence to best practice protocol very seriously				
9. My employer organizes formal training to support best practices				

officer, two nursing officers and a PhD midwifery student. A copy of the draft questionnaire and assessment framework was emailed to each of the members of the panel. To ensure anonymity, the members were blinded from one another while the completed assessment framework was sent individually by email within 2 weeks. The first draft was sent to a 9-member panel of nurses and doctors working in public maternity care in Nigeria for assessing (external) face validity of the questionnaire. Disagreement on an item was revised with a consultation with the literature and the theoretical framework. None of the reviewer was directly known to the researchers, and review was done voluntarily without any incentive.

2.2. Step 3: Administration and Assumption Test

2.2.1. Ethical Approval. The details of the research questions, objectives, data to be collected, methods of data collection and analysis were submitted with an introduction letter to the nurses’ association and University of Huddersfield. Participants’ rights were guaranteed, and approval was granted through emails received by the nurses’ association and the University of Huddersfield’s ethical approval

committee. This study received ethical approval on 8 October 2020 from the School Research Ethics and Integrity Committee of the University of Huddersfield with approval reference OMOLADE (PhD)-SREIC PGR Panel Application-SREIC/2020/088–Outcome.

Also, on 01 October 2020, the National University Nursing Student Association gave an approval letter for the survey link to be shared with the qualified nurses and midwives completing their top-up degree. The study adheres to all ethical standards including anonymity, confidentiality and voluntary participation or withdrawal.

2.2.2. Population and Sample Size. Participants for this study were nurses and midwives in the Nigeria public hospitals. The design of this survey formed part of a cross-sectional study that examined the relationship between the culture of safety in public maternities and treatment of primary postpartum haemorrhage. Both paper and virtual link to the nine-item survey were made accessible to a cohort of qualified nurses and midwives who have completed or about to complete a (top-up) nursing and midwifery degree in Southwest Nigeria. Determining the sample size for the Rasch analysis is based on Linacre’s suggestion that

a minimum of 1:10 (item-to-person ratio) is desirable but any questionnaire, notwithstanding the number of items, can be correctly analysed with 500 respondents [20, 27, 28]. On this estimation, a sample size of 90 responses would have been enough but 967 participants far exceeded the requirement for the analysis.

**2.2.3. Study Area.** Respondents to the survey were nurses and midwives working within the same catchment area of maternity settings in Southwest Nigeria. Nigeria's healthcare system is structured into three levels: primary, secondary and tertiary care, each managed by different government tiers. The primary health care (PHC) operates at the community level and is the first point of contact for patients. The secondary care is a referral point for PHC facilities, handling more complex health issues, while the tertiary facilities offer specialized medical services, advanced treatments, teaching and training services.

**2.2.4. Data Analysis and Measures.** Both the descriptive statistics of the participants and psychometric assumptions were assessed under Rasch techniques. Building on suggestions and guidelines on rating scale design [18, 20, 21, 23, 29, 30], we evaluated the culture of patient safety scale for six assumptions of objective measurement including item difficulty, dimensionality, category functioning, fit statistics, local independence, reliability and scale validity using Winsteps Version 5.0.0.0 [26, 28]. Applying Rasch measurement model is the cornerstone of developing a high-quality (objective) rating scale because it ensures unidimensionality, invariance and proper category functioning, leading to reliable, valid and meaningful assessments of the intended construct [16, 26]. Correspondingly, the Rasch partial credit model (PCM) [17, 18, 21] dominated the psychometric assessment conducted in this study and only Cronbach's alpha value was estimated as a method from the CTT. PCM equation [18] is presented as follows:

$$\ln(P_{ni}(x_i = k - 1)P_{ni}(x_i = k)) = \theta_n - \delta_i - \tau_{ik}, \quad (1)$$

where  $(P_{ni}(x_i = k))$  is the probability of person ( $n$ ) scoring in category ( $k$ ) on item ( $i$ ).  $(\theta_n)$  is the ability of person ( $n$ ).  $(\delta_i)$  is the difficulty of item ( $i$ ).  $(\tau_{ik})$  is the threshold parameter for category ( $k$ ) on item ( $i$ ).

PCM is an advancement of the dichotomous Rasch technique for analysing items thus suitable for Likert-type rating scales [18]. The PCM allows each item to have its own unique set of thresholds, which represent the points at which the probability of endorsing a higher category surpasses that of a lower one [24, 26]. It helps in detecting disordered thresholds, which indicate that respondents do not perceive the categories in a consistent, ordinal manner. By doing so, it ensures that the scale measures a single underlying trait or construct, enhancing the validity and reliability of the questionnaire [17, 20].

**2.2.4.1. Demographic Data.** To obtain context-informed perspective from participants, the included information

on gender, place of work, present grade level, qualifications and availability of best practice guideline on their units. Where appropriate, continuous and categorical measures (frequencies and percentages) were used to present participants' sociodemographic data.

**2.2.4.2. Item Difficulty Assessment.** This essential psychometric assessment is conducted using Wright maps to show a linear interval relationship of the items on the rating scale [20]. Evaluating item difficulty ensured that each item accurately reflected the latent culture of safety regardless of participants' ability. This assessment helps in enhancing instrument's reliability and validity by identifying items that may be measurement redundant. Moreover, it allows for the creation of balanced surveys that can differentiate between varying levels of ability, ensuring precise and meaningful measurement outcomes. A key contribution of Wright map to psychometric assessment is that identical locations of the average measures (mean score) for the items' difficulty level ( $M$ ) and respondents' ability ( $M$ ) on the logit scale indicate the appropriateness of the scale (not too easy nor too difficult) to the target respondents [20, 26].

**2.2.4.3. Dimensionality.** Assessing dimensionality (Rasch residual) is crucial in Rasch analysis to ensure that a measurement instrument accurately reflects a single underlying trait or construct [31]. This diagnostic step helps to identify whether items on a test or survey measure the same latent variable or an unintended construct. The assumption that the culture of safety scale is a single dimension was examined through a principal component analysis of the residuals (PCAR) frequently applied in Rasch analysis. For acceptable dimensionality of the scale, first residual factor must explain at least 50% of the variance and the eigenvalues of the residuals should be less than 2.0 [29].

**2.2.4.4. Category Functioning.** Assessing the category functioning of a rating scale involved several key parameters such as thresholds between categories, category frequency and fit statistics, which identify any misfitting responses that could distort the measurement [18, 31, 32]. The aim of this assessment is to objectively evaluate the consistency of progression of the category order as proposed and used by the respondents.

**2.2.4.5. Local Independence.** Evaluating the local independence of items on a rating scale is crucial in this study to ensure that each item measures the intended construct without being influenced by responses to other items. The statistics used in this assessment help identify pairs of items that may be too closely related indicating potential redundancy or dependency through residual correlations [18]. A rule of thumb is that interitem correlations should be very low (far less than one), suggesting that the items independently contribute to the measurement of the construct under investigation [18].

**2.2.4.6. Item Validity and Reliability.** Applying the Rasch PCM, validity of items on culture of safety scale were

assessed for fitting with Rasch measurement model and linear order of items on the logit scale. Item fit statistics, such as infit (information-weighted fit) and outfit (outlier-sensitive fit) mean-square values, were assessed to identify misfitting items. Acceptable values for a good fit must range from 0.5 to 1.5 for item validity [18]. Concurrently, reliability assessment must yield a score of 0.7 (or even higher) to indicate high item difficulty variance, adequate sample size and finally, good internal consistency [18, 26].

Table 2 outlines the psychometric properties proposed for assessment in this study.

### 3. Results

In this section are presented the descriptive statistics of the participants, outcome of the assessment of face validity, Wright map analysis (logit scale), category test, unidimensionality, local independence, item fit statistics, reliability and validity outcomes.

**3.1. Descriptive Statistics.** In Table 3 are the sociodemographic descriptors of the 967 participants including gender, place of work, grade level, academic qualifications and availability of best practice policy in the hospitals. For each of the expected demographics for the 967 participants, 25% were missing. Overall, 85.6% of the participants were female and all the three levels of Nigerian public maternity settings were adequately represented. Less than 2% of respondents have master degree education against the majority that are either specialist or have a nursing degree.

**3.2. Step Two: Feedback From the External Face Validity Assessment.** The result for the external validation is compiled from the comments and observation reported by the nine-member panel including remarks on clarity, simplicity, relevance and general presentation of the culture of patient safety scale in Table 4.

**3.3. Item Difficulty (Wright Map) Analysis.** Wright map (Figure 2) analysis is an innovative technique that engages probabilistic model to present complicated rating scale (survey) items of culture of patient safety on a linear interval logit scale. The item measures are rescaled on a linear logit scale 0 to 100 logits displaying the level of difficulty (agreeability) of the nine items on the culture of patient safety, and the emphasis of Wright map is the distribution of the items on the interval logit scale. The first remarkable feature of this map is the gap created by missing items on the scale from 0 to 35 logits, 46 to 55 logits and 66.5 to 100 logits. To the contrary, points 36 to 45 logits and 56 to 65 logits are clustered by five items (Q4, Q7, Q6, Q9 and Q5) and four items, respectively (Q8, Q3, Q2 and Q1). The average difficulty level between the five items measuring at 36 to 45 logits is 2.7 logits and the same for the four items measuring at 46 to 55 logits. The average difficulty of this scale and participants' ability is marginally different by 1.7 logit, implying the scale is neither too easy nor too difficult for

participants. Item Q1 is the least agreeable (most difficult), while item Q4 is the easiest for participants to agree and no item indicating the average measure of the scale. Two items Q6 and Q9 measure at equal level of difficulty of 42.4 logits, suggesting one of the item measures is redundant. A key observation of this scale is that participants' scoring less than 56 logit on the scale is less likely to agree to four items (Q1, Q2, Q3 and Q8).

**3.4. Unidimensionality and Local Independence.** The unidimensional characteristic of this scale was assessed and confirmed as the eigenvalue of the first unexplained contrast is 1.60 (less than 2) and the model accounted for 54% of the explained variance. Also, items were tested for local dependency and the evidence on Table 5 shows the largest standardized residual correlations showed a negative correlation for all the 9 items. This evidence shows that there is no interitem dependency for the 9 items implying no interitem interference.

**3.5. Category Function Test.** In Table 6 named category function, there is a good distribution of participants across all the categories with at least more than ten responses per category. In line with the fundamental principle of measurement that Category 1 suggests less measure and Category 4 suggests the highest possible measure, and the Andrich threshold monotonically increased from none to 18.84. This evidence shows that the category functioning of the reinforcement scale is good. The category functioning is further illustrated on the category probability curves (see Figure 3), and the point of intersection between the adjacent category represents the Andrich thresholds with a good increase.

Scale reliability assessed under Rasch model is 1 and Cronbach's alpha value is 0.87, indicating the items on the questionnaire have excellent internal consistency.

**3.6. Item Fit With Rasch Measurement Theory.** In Table 7, the lowest infit MNSQ on the scale is 0.77 and the highest is 1.12, while the outfit values range between 0.75 and 1.13, indicating that all the items are within acceptable fit with the Rasch model (acceptable values range from 0.5 to 1.5).

## 4. Discussion

This study engaged rigorous theoretical and statistical steps in developing and validating the culture of patient safety scale (often used in public maternity settings in Nigeria) for two main reasons. First is to objectively assess and display the psychometric properties of the scale so that clinicians are reassured of the quality of the precision and accuracy of the scale as a valid measuring instrument. Second is to advance objective measurement theory into the domain of designing patient safety scales where the severely limited traditional methods (CTT) have widely dominated. Before the invention of objective measurement techniques, controversies surround the psychometric properties of majority of the

TABLE 2: Diagnostic indicators of the psychometric assessment.

Diagnostic indicators	Suggested standards
Face validity	Agreement among panel of reviewers
Wright map analysis	Linear arrangement on the logit scale
Category function	Monotonously increasing thresholds and measures across the categories
Scale dimensionality	Assessed using the principal component analysis of residuals (PCAR) with eigenvalue of the first unexplained contrast less than 2 and more than 50% of explained variance by the Rasch measures
Local independence	Interitem correlations of residuals must be (low) less than 0.5
Item fit statistics	Estimated from infit and outfit mean-square values (0.5 to 1.50)
Scale reliability	0.7 and above indicates good scale reliability

TABLE 3: Participants' characteristics.

Demographics	Frequency	Valid percent
Gender	Male	10.3
	Female	85.6
	Prefer not to say	3.8
	Total	100.0
	Missing	225
	Total	967
Place of work	Primary healthcare centres	25.1
	General hospitals	38.0
	Teaching hospitals	36.9
	Valid total	100.0
	Missing	225
	Total	967
Grade level	Assistant/chief nursing officer	21.0
	Principal nursing officer	13.2
	Senior nursing officer	14.4
	Nursing officer 1&2	51.4
	Missing	100
	Total	967
Qualifications	Master's degree and above	1.9
	Nursing degree	32.9
	Post-basic nursing cert	14.3
	RN&RM	28.9
	RN only	28.9
	RM only	14.7
	Missing	100
	Total	967
Availability of safety policy in my hospital	Yes	62.7
	No	37.3
	Missing	100
	Total	967

TABLE 4: Feedback from panel of reviewers.

Items	Yes	No	Remarks (why NO)
1. The questionnaire is easy to understand	9	0	None
2. The general presentation of the questionnaire is easy to follow	9	0	None
3. The content of the questionnaire is relevant to patient safety	9	0	'I find this scale relevant to patient safety but no item on clinician's adherence with best practices'
4. The words used in each item are simple enough to read	9	0	None
5. The instructions provided on completing the questionnaire are very clear	9	0	None
6. Average time to complete the questionnaire is okay	9	0	None

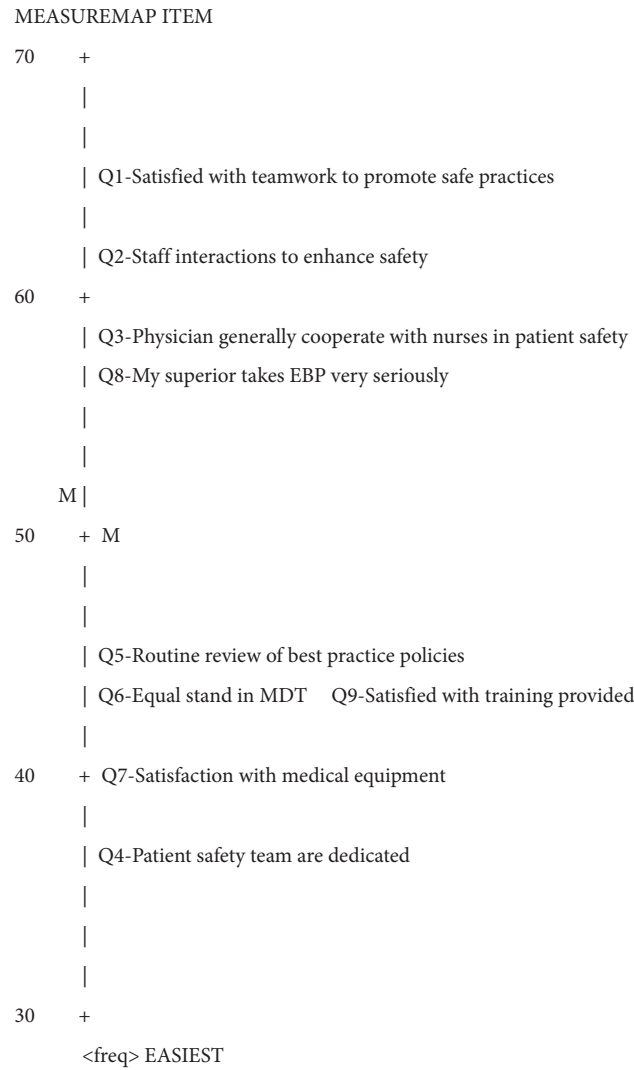


FIGURE 2: Item difficulty spread assessment.

TABLE 5: Result of local independence of items.

Correlations	Item entry number	Item entry number
-0.24	Q2	Q9
-0.21	Q3	Q4
-0.21	Q3	Q5
-0.20	Q3	Q7
-0.20	Q1	Q9
-0.14	Q4	Q9

TABLE 6: Category function test.

Category label	Category score	Observed count	Infit MNSQ	Outfit MNSQ	Andrich threshold	Category measure
1	1	2071	1.05	1.04	None	-28.38
2	2	4227	0.95	0.98	-20.79	-9.52
3	3	1693	0.94	0.95	-1.95	10.46
4	4	234	1.05	1.11	18.84	26.65

‘validated’ culture of safety scales due to lack of evidence of objective functions of the scales. However, the methodologies here mirror a departure from the traditional approach

because the results of the statistical analyses provide evidence-based rationale for fine-tuning each item on the rating scale in line with objective measurement theory. The



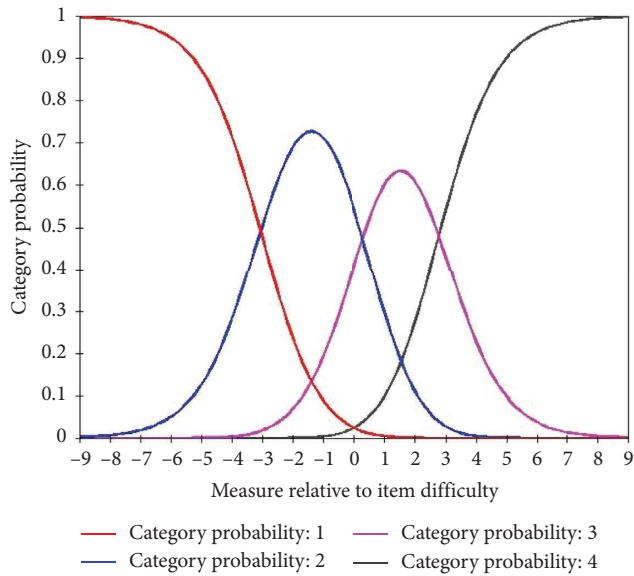


FIGURE 3: Category probability curve for Q1 (and the same for all the items).

TABLE 7: Fit statistics of the scale to Rasch measurement model.

Items	Infit (MNSQ)	Outfit (MNSQ)
Q6	1.12	1.13
Q1	1.04	1.09
Q4	1.06	1.09
Q7	1.05	1.08
Q9	1.02	1.04
Q3	0.98	1.01
Q5	0.91	0.93
Q2	0.91	0.91
Q8	0.77	0.75

psychometric properties assessed and presented in the result section include the category functioning, dimensionality, scale reliability, validity and the Cronbach's alpha. This discussion will explore the broader impact of these findings, potential limitations and directions for future research to further enhance the application of Rasch methods in psychometric assessments in health and human sciences.

After analysing 967 responses to the culture of patient safety under objective measurement techniques, the results showed that the rating scale has the following:

- Strong face validity as shown by the unanimous agreement of the expert review panel
- High internal consistency demonstrated by a strong Cronbach's alpha value of 0.87
- Strong evidence of unidimensional construct validity tested under PCAR
- Linearly related items displayed on the Wright map analysis on the logit scale
- Excellent category functioning that showed the even spread of the category order including evidence of monotonous increase from the lower order to higher category order

- Negative interitem correlations showing convincing evidence that each item can elicit independent responses from participants without interfering with another item on the scale
- Strong evidence of content validity and reliability assessed using Rasch techniques
- Acceptable fitting of the items to objective measurement model showed with the infit and outfit statistics

In developing various scales on culture of safe practices in healthcare settings, remarkable disparities exist in the methodologies [3, 7, 15] but no scale, to the best of our knowledge, has engaged objective measurement theory to design a scale of culture of patient safety. Overall, the results here display rigorous approach and assessment methods under Rasch theory to ensure the culture of patient safety scale present convincing psychometric properties. By applying Rasch model, a quality assurance of the culture of safety scale as a precision measuring tool was demonstrated.

The results from theoretical input and assessment of face validity by the review panel are considered the preliminary steps which feed into the evidence for construct validity of this current scale. Except for literature review, inputs from professionals acquainted with the practice areas are recurrently missing in previous studies. Meanwhile, expert review panel plays a crucial role in Rasch techniques to determine clarity and relevance of the items and predict the linear relationship among items on the basis of perceived level of difficulty [20]. Linear relationships of items are not proven in the CTT approach but frequently presented using Wright map analysis in Rasch techniques.

In combination, both reliability and validity assessment of this scale indicate excellent psychometric function yet proponents of Rasch techniques emphasize the revolutionary role of Wright map in developing high-quality surveys and questionnaires. Correspondingly, attention was focused on the Wright map analysis to identify important gaps on the culture of patient safety scale requiring exploration for improved measurement function. The items [1–9] presented on a linear logit rating of 0 to 100 logits ought to display indicators parallel to the grades within acceptable or uniform consistency. However, logit grades at points 0 to 30, 46 to 55 and 66 to 100 logits missed out items, which may suggest areas of patient safety culture that were not captured by the scale. Correspondingly, the implication is to explore more literature on culture of patient safety to identify items suitable to occupy the missed areas.

Also, it is interesting to note the pattern of discordant relationship between items that delved into perceived dedication of patient safety team and teamwork to enhance safety (Q1 and Q4). An inference from this result is that despite the dedication of the safety team, team-working is perceived as a barrier to culture of safety since the item has the lowest level of agreeability. Furthermore, items 6 and 9 measure at equal levels of 43.5, indicating one of the items is redundant and may be revised to lesser or higher levels of difficulty. In line with the criteria of good category functioning [31] as essential psychometric property, the result of this study displayed



excellent function of the category order. A combination of diagnostic parameters assessed to confirm the objective function of the culture of safety scale included the following: [1] evidence of categories having more than ten responses each, [2] monotonous increase of average measures and thresholds from lower to higher category, [3] infit and outfit mean-square values from 0.5 to 1.5 and [4] the gaps between each category measure higher than 1.4 logits but less than 5.0 logits.

Alternatively, in many other culture of patient safety scales, objective psychometric assumptions such as item difficulty and category functioning were not evaluated but Cronbach's alpha and confirmatory factor analysis (CFA) were conducted [15]. Even though an item on a rating scale may be perceived theoretically relevant to patient safety, methods often cited for assessing dimensionality under CTT include CFA estimators such as diagonally weighted least-squares (WLSMV), model fit, comparative, Tucker–Lewis's index (TLI), root-mean-square error of approximation (RMSEA) and nonsignificant chi-square [15]. Nonetheless, the unidimensionality of a scale may be assessed through the PCAR under Rasch analysis, CFA under CTT or the combination of both methods [30, 31]. The practical usefulness for combining both techniques are limited; therefore, PCAR is the dominant technique to prove the dimensionality of a survey under Rasch theory of designing [18].

The moderate respondent population ( $N=967$ ), involvement of the three levels of Nigeria public settings and rigorous assessments of the psychometric properties under Rasch techniques constitute some of the major strengths of the current study. However, some limitations are still inherent in this study. First is that despite applying objective measurement theory in this study, the scale development is inevitably open to participants' biases including memory recall, social desires and undisclosed motivations to answer the survey questions. Meanwhile, the areas of missing items on the scale (Wright map) can be addressed by incorporating qualitative methods for understanding the subjective experiences and perceptions of healthcare professionals and patients which quantitative data alone cannot unveil. Additionally, we propose using qualitative or mixed methodology to investigate culture of patient safety so that a contextualized understanding of safety practices and culturally sensitive interventions and policies may evolve.

Another limitation is that all the respondents are nurses and midwives from the same catchment area of public maternity settings in Nigeria despite a moderate sample of 967 respondents. An unplanned but useful finding in this study is that the respondents preferred the online survey to paper copies, which may be a valid antecedent for future researchers in making methodological decisions. Correspondingly, we recommend to future studies to expand the scope of studies on patient safety in Nigeria to include hospital in other states or regions and include diverse population of healthcare workers (such as doctors, laboratory scientists, pharmacists, management staff and non-technical staff). Finally, the data collection for this study only lasted about 8 weeks suggesting the cross-sectional design of the study. Even though objective measurement theory

ensures measures are independent of population, longitudinal test-retest assessment of the psychometric properties is encouraged.

## 5. Implications

The findings of this current study provide valuable insights for clinicians, researchers and policymakers, guiding improvements in patient safety practices especially in the Nigeria maternity settings. The study provides clinicians with a reliable tool to assess and improve hospital safety culture. Applying this tool may lead to better identification of safety issues, more effective interventions and, ultimately, enhanced patient care. Clinicians can use the scale to foster a more open and communicative environment, encouraging reporting and discussion of safety concerns. In line, policymakers can leverage the findings to develop and implement evidence-based policies aimed at improving patient safety. Additionally, applying the scale can help in setting benchmarks and standards for safety culture, thereby guiding regulatory and accreditation processes. For researchers, the study offers a validated instrument for measuring patient safety culture and the applied Rasch techniques simplified the processes in developing surveys as rating scales of health and patient (self-reported) outcomes. In conclusion, engaging the objective measurement theory in this study presents the psychometric properties of culture of safety scales as high-quality and valid rating scale that can help determine targeted interventions fostering a safer environment for patients and staff. These findings support ongoing efforts to enhance patient safety in the maternity settings in Nigeria and inform future research and policy development on culture of safety.

## Data Availability Statement

Data are available on request.

## Ethics Statement

Approval received from the University and participants.

## Consent

The authors have nothing to report.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Funding

No funding was received for this research.

## References

- [1] O. K. Omolade and J. Stephenson, "Best Rating Scale Design Theory: Implications for Developing Questionnaires in Nursing and Health Sciences," *Journal of Nursing Practices and Research* 3, no. 4 (2023): 25, <https://doi.org/10.53964/jmnpr.2023025>.

- [2] M. H. DiCuccio, "The Relationship Between Patient Safety Culture and Patient Outcomes: A Systematic Review," *Journal of Patient Safety* 11, no. 3 (2015): 135–142, <https://doi.org/10.1097/pts.0000000000000058>.
- [3] K. Churruca, L. A. Ellis, C. Pomare, et al., "Dimensions of Safety Culture: A Systematic Review of Quantitative, Qualitative and Mixed Methods for Assessing Safety Culture in Hospitals," *BMJ Open* 11, no. 7 (2021): e043982, <https://doi.org/10.1136/bmjopen-2020-043982>.
- [4] B. M. Melnyk, "Culture Eats Strategy Every Time: What Works in Building and Sustaining an Evidence-Based Practice Culture in Healthcare Systems," *Worldviews on Evidence-Based Nursing* 13, no. 2 (2016): 99–101, <https://doi.org/10.1111/wvn.12161>.
- [5] V. Raftopoulos, N. Savva, and M. Papadopoulou, "Safety Culture in the Maternity Units: A Census Survey Using the Safety Attitudes Questionnaire," *BMC Health Services Research* 11, no. 1 (2011): 238, <https://doi.org/10.1186/1472-6963-11-238>.
- [6] S. J. Weaver, L. H. Lubomksi, R. F. Wilson, E. R. Pfoh, K. A. Martinez, and S. M. Dy, "Promoting a Culture of Safety as a Patient Safety Strategy: A Systematic Review," *Annals of Internal Medicine* 158, no. 5\_Part\_2 (2013): 369–374, <https://doi.org/10.7326/0003-4819-158-5-201303051-00002>.
- [7] C. T. Reis, S. G. Paiva, and P. Sousa, "The Patient Safety Culture: A Systematic Review by Characteristics of Hospital Survey on Patient Safety Culture Dimensions," *International Journal for Quality in Health Care* 30, no. 9 (2018): 660–677, <https://doi.org/10.1093/intqhc/mzy080>.
- [8] S. D. Wami, A. F. Demssie, M. M. Wassie, and A. N. Ahmed, "Patient Safety Culture and Associated Factors: A Quantitative and Qualitative Study of Healthcare Workers' View in Jimma Zone Hospitals, Southwest Ethiopia," *BMC Health Services Research* 16 (2016): 495–510, <https://doi.org/10.1186/s12913-016-1757-z>.
- [9] M. Hedsköld, K. Pukk-Härenstam, E. Berg, et al., "Psychometric Properties of the Hospital Survey on Patient Safety Culture, HSOPSC, Applied on a Large Swedish Health Care Sample," *BMC Health Services Research* 13 (2013): 1–11.
- [10] N. Alquwez, J. P. Cruz, A. M. Almoghairi, et al., "Nurses' Perceptions of Patient Safety Culture in Three Hospitals in Saudi Arabia," *Journal of Nursing Scholarship* 50, no. 4 (2018): 422–431, <https://doi.org/10.1111/jnu.12394>.
- [11] J. Macallan, J. Sutcliffe, and S. Lomax, "Human Factors in Anaesthetic Practice Part I: Facts and Fallacies," *BJA Education* 23, no. 10 (2023): 398–405, <https://doi.org/10.1016/j.bjae.2023.06.002>.
- [12] D. Phipps, G. H. Meakin, P. C. Beatty, C. Nsoedo, and D. Parker, "Human Factors in Anaesthetic Practice: Insights from a Task Analysis," *British Journal of Anaesthesia* 100, no. 3 (2008): 333–343, <https://doi.org/10.1093/bja/aem392>.
- [13] C. Jones, J. Fawker-Corbett, P. Groom, B. Morton, C. Lister, and S. J. Mercer, "Human Factors in Preventing Complications in Anaesthesia: A Systematic Review," *Anaesthesia* 73, no. S1 (2018): 12–24, <https://doi.org/10.1111/anae.14136>.
- [14] L. Mitchell and R. Flin, *Safer Surgery: Analysing Behaviour in the Operating Theatre* (CRC Press, 2017).
- [15] D. Bartonickova, D. Kalankova, and K. Ziakova, "How to Measure Patient Safety Culture? A Literature Review of Instruments," *Acta Medica Martiniana* 21, no. 2 (2021): 69–79, <https://doi.org/10.2478/acm-2021-0010>.
- [16] O. Kolawole Omolade and J. Stephenson, "Best Practices in Reporting Analyses of Questionnaires as Objective Rating Scales of Variable Measures," *Nurse Researcher* 32, no. 1 (2024): 10–18, <https://doi.org/10.7748/nr.2023.e1903>.
- [17] P. Oon and X. Fan, "Rasch Analysis for Psychometric Improvement of Science Attitude Rating Scales," *International Journal of Science Education* 39, no. 6 (2017): 683–700, <https://doi.org/10.1080/09500693.2017.1299951>.
- [18] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (Psychology Press, 2013).
- [19] E. B. Tunç, "A Review of Measurement Tools Developed and Adapted Based on the Rasch Model," *İnsan ve Sosyal Bilimler Dergisi* 6, no. 2 (2023): 249–275, <https://doi.org/10.53048/johass.1369336>.
- [20] W. J. Boone and A. Noltemeyer, "Rasch Analysis: A Primer for School Psychology Researchers and Practitioners," *Cogent Education* 4, no. 1 (2017): 1416898, <https://doi.org/10.1080/2331186x.2017.1416898>.
- [21] W. J. Boone, "Rasch Analysis for Instrument Development: Why, When, and How?" *CBE-Life Sciences Education* 15, no. 4 (2016): rm4, <https://doi.org/10.1187/cbe.16-04-0148>.
- [22] W. J. Boone, J. R. Staver, and M. S. Yale, *Rasch Analysis in the Human Sciences* (Springer, 2013).
- [23] W. Ling Lee, K. Chinna, and B. Sumintono, "Psychometrics Assessment of HeartQoL Questionnaire: A Rasch Analysis," *European Journal of Preventive Cardiology* 28, no. 12 (2021): e1–e5, <https://doi.org/10.1177/2047487320902322>.
- [24] C. Van Zile-Tamsen, "Using Rasch Analysis to Inform Rating Scale Development," *Research in Higher Education* 58, no. 8 (2017): 922–933, <https://doi.org/10.1007/s11162-017-9448-0>.
- [25] A. E. Patelarou, A. Laliotis, H. Brokalaki, J. Petrakis, V. Dafermos, and E. Koukia, "Readiness for and Predictors of Evidence Base Practice in Greek Healthcare Settings," *Applied Nursing Research* 35 (2017): 64–70, <https://doi.org/10.1016/j.apnr.2017.02.003>.
- [26] W. J. Boone and J. R. Staver, *Advances in Rasch Analyses in the Human Sciences* (Springer, 2020).
- [27] J. M. Linacre, "Sample Size and Item Calibration Stability," *Rasch measurement transactions* 7 (1994): 328.
- [28] J. M. Linacre and B. D. Wright, "Winsteps," (2000), <http://www.winsteps.com/index.htm>.
- [29] C. Combrinck, *The Use of Rasch Measurement Theory to Address Measurement and Analysis Challenges in Social Science Research* (South Africa: University of Pretoria, 2018).
- [30] N. Sakib, A. K. M. I. Bhuiyan, S. Hossain, et al., "Psychometric Validation of the Bangla Fear of COVID-19 Scale: Confirmatory Factor Analysis and Rasch Analysis," *International Journal of Mental Health and Addiction* 20, no. 5 (2022): 2623–2634, <https://doi.org/10.1007/s11469-020-00289-x>.
- [31] O. K. Omolade, J. Stephenson, P. Simkhada, and A. Keely, "Is This a Good Questionnaire? Dimensionality and Category Functioning of Questionnaires Used in Nursing Research," *Nurse Researcher* 32, no. 3 (2024).
- [32] J. M. Linacre, "Optimizing Rating Scale Category Effectiveness," *Journal of Applied Measurement* 3, no. 1 (2002): 85–106.