Indonesian Journal of Education & Mathematical Science
Vol. 7, No. 2, Mei 2026, pp. 161~174
ISSN: 2715-985X, DOI: 10.30596/ijems.v7i2.29464

Detecting Differential Item Functioning in the Measurement of Ethnomathematical Knowledge among Preservice Teachers: A Rasch Analysis.

Ilham Falani[1], Yelli Ramalisa[2], Nurfadilah[3], Inda Lestari[4], Destri Yaldi[5]
[1,2,3,4,5]Department of Mathematics Education, Universitas Jambi, Jambi, Indonesia
[1]ilhamfalani@unja.ac.id, [2]yelliramalisa@unja.ac.id, [3]nurfadillah@unja.ac.id, [4]indalestari@unja.ac.id, [5]destriyaldi@unja.ac.id,

ABSTRACT

This study examines Differential Item Functioning (DIF) in an instrument developed to measure preservice mathematics teachers' ethnomathematical knowledge in Jambi. Anchored in the Rasch measurement framework, this study employs a quantitative ex post facto design to evaluate item-level measurement invariance of an ethnomathematics assessment instrument. Rasch-based procedures: item and person parameter estimation, fit statistics, and DIF analyses were employed to assess item equivalence across gender, educational background, and teaching experience, and to quantify the magnitude of any detected DIF. Data from 150 preservice teachers were analyzed using Rasch estimation and Mantel–Haenszel DIF on a culturally contextualized ethnomathematics instrument. Descriptive analyses indicated strong psychometric properties of the instrument, with person reliability = 0.81, item reliability = 0.97, and fit statistics falling within acceptable ranges. DIF analyses revealed several items with statistically significant DIF predominantly across gender, while the majority of items demonstrated stable functioning across the examined participant subgroups. Graphical DIF plots revealed heterogeneous item-response patterns in the ethnomathematics instrument, highlighting the need for culturally responsive item revision to reduce bias. Findings demonstrate the Rasch model's effectiveness in detecting item-level bias in instruments measuring preservice teachers' ethnomathematical knowledge and substantiate targeted refinement of measurement instruments.

Keywords: DIF, Ethnomathematics, Rasch Model

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:
Ilham Falani,
Mathematics Education Program Study,
Universitas Jambi
Jl. Raya Jambi - Muara Bulian, KM 15, Mendalo Darat, Muaro Jambi, Jambi, 36361, Indonesia
ilhamfalani@unja.com

1. INTRODUCTION

Ethnomathematics has emerged as a pivotal discipline that bridges mathematics and cultural heritage, contributing significantly to the preservation and revitalization of indigenous knowledge systems (Rosa et al., 2016; Rosa & Orey, 2021). Defined as the study of mathematical practices embedded within cultural contexts, ethnomathematics explores how traditional societies conceptualize and utilize mathematical ideas in everyday activities such as measurement systems, architectural patterns, handicrafts, and ceremonial traditions (Bahar et al., 2023; Ergene et al., 2020; Putra & Mahmudah, 2021; Salam et al., 2022; Sunzuma & Maharaj, 2019). This culturally grounded perspective not only enhances our understanding of mathematics as a human endeavor but also serves as a means of safeguarding intangible cultural heritage through educational integration (Ramalisa et al., 2023).

Teachers serve as pivotal facilitators in bridging the gap between abstract mathematical concepts and cultural realities. By integrating ethnomathematical principles into curricula and instructional strategies, educators can transform mathematical learning into a more contextualized, meaningful, and engaging experience for students (Ergene et al., 2020; Mania & Alam, 2021; Prahmana, 2022). This integration involves leveraging local narratives, artifacts, and community practices, which not only enhances pedagogical effectiveness but also fosters an inclusive environment that celebrates cultural diversity a hallmark of contemporary multicultural education (Atmaja, 2023; Mosimege & Egara, 2023; Nur et al., 2020; Putra & Mahmudah, 2021). Consequently, it is imperative for preservice teachers to develop a robust grasp of both the theoretical foundations and practical applications of ethnomathematics. Ensuring that their knowledge measures are free from bias is a critical step in preparing them to implement culturally responsive education effectively in a globalized world (Mania & Alam, 2021; Ramalisa et al., 2023).

Given the deeply contextual nature of ethnomathematics, developing a standardized measure of this knowledge presents unique psychometric challenges. Since these assessments inherently draw upon cultural narratives and local practices, there is a heightened risk that certain items may inadvertently favor specific demographic or cultural groups over others (Desai et al., 2021). To address this concern and ensure the validity of the findings, Differential Item Functioning (DIF) analysis becomes an indispensable methodological approach. DIF functions by identifying test items that exhibit different statistical properties across diverse subgroups, even when those individuals possess the same underlying ability level (Skaggs & Williams, 2023). In the realm of ethnomathematics, utilizing DIF is essential not merely as a statistical check, but as a commitment to assessment equity; it guarantees that the instrument remains culturally neutral and accurately reflects the preservice teachers' professional competence rather than their cultural or linguistic familiarity (Nicol et al., 2024).

Several methods can be employed to detect DIF in assessment instruments, one of the most widely recognized being the Rasch Model (Ramalisa et al., 2023). The Rasch Model offers distinct advantages for DIF analysis, making it an effective and rigorous approach for ensuring fairness in the measurement of ethnomathematical knowledge among prospective mathematics teachers. One of the model's key strengths lies in its principle of measurement invariance, which enables the estimation of respondent abilities independently of the specific set of items administered (Falani et al., 2022). This allows for more accurate detection of item-level bias across groups.

Moreover, the Rasch Model maintains parameter separability, meaning that item characteristics are estimated independently from the distribution of respondent abilities (Keliat et al., 2023). This property enhances the objectivity of the analysis by ensuring that sample-specific factors do not influence item functioning. Compared to traditional (classical) approaches, the Rasch Model demonstrates greater sensitivity in detecting DIF, thereby allowing for more precise identification of differential item performance across subgroups.

Additionally, the application of the Rasch Model contributes significantly to both the validity and reliability of the measurement instrument (Falani et al., 2022; Keliat et al., 2023). By supporting the development of culturally fair assessment tools, it ensures that prospective teachers are evaluated in a manner that truly reflects their understanding of ethnomathematics free from cultural or contextual bias. Its ability to produce linear, interval-scaled measures also facilitates clearer interpretation of results, further reinforcing its value as a psychometrically sound approach. For these reasons, the Rasch Model has been selected as the primary method for DIF analysis in this study.

A preliminary bibliometric analysis using VOSviewer software (Falani et al., 2022; Keliat et al., 2023) highlights a significant research gap: the absence of DIF analysis in ethnomathematical instruments specifically tailored to the Jambi cultural context. Jambi possesses a distinct cultural identity characterized by a rich tapestry of Malay-Islamic influences, manifested in intricate geometric patterns of Batik Jambi (such as the Angso Duo and Kapal Sanggat motifs) and the sophisticated architectural symmetry of the Rumah Larik or Panggung houses. These cultural artifacts are not merely aesthetic; they embody complex ethnomathematical concepts, including transformational geometry, tessellation, and spatial logic. Despite this potential, no existing studies have examined how these specific local nuances might influence the functioning of assessment items.

This research gap warrants serious attention, as an assessment that fails to account for these unique cultural markers may result in item bias, thereby obscuring the true ethnomathematical competence of preservice teachers. Since prospective teachers are the primary agents in bridging local wisdom with formal mathematics, the development of contextually grounded and psychometrically valid tools is vital. By ensuring that instruments are free from DIF relative to Jambi's specific cultural characteristics, this study not only promotes assessment equity but also serves as a model for safeguarding regional cultural heritage through rigorous, culturally responsive educational measurement.

A systematic bibliometric analysis of 532 scholarly articles indexed in global databases between 2020 and 2025, as visualized in Figure 1, provides empirical evidence of a significant research lacuna concerning the application of Differential Item Functioning (DIF) analysis within ethnomathematical contexts. The visualization revealed a lack of structural co-occurrence and node connections between the term "ethnomathematics" and other critical keywords associated with psychometric bias or teacher competency (Donthu et al., 2021). This isolation indicates that while ethnomathematics has gained

Indonesian Journal of Education & Mathematical Science
Vol. 7, No. 2, Mei 2026, pp. 161~174
ISSN: 2715-985X, DOI: 10.30596/ijems.v7i2.29464

traction as a pedagogical concept, its rigorous psychometric evaluation, specifically regarding the ethnomathematical knowledge of prospective teachers in localized contexts like Jambi, remains at a nascent stage of academic inquiry. Compared to global trends in mathematical literacy assessments which increasingly utilize IRT-based diagnostics (Masters, 2020), the absence of DIF-related nodes in this dataset underscores a critical 'measurement void.' This study, therefore, addresses this gap by synthesizing cultural mathematics with advanced psychometric modeling to ensure a fair and non-biased evaluation framework (Alkharusi, 2020; Retnawati et al., 2018).

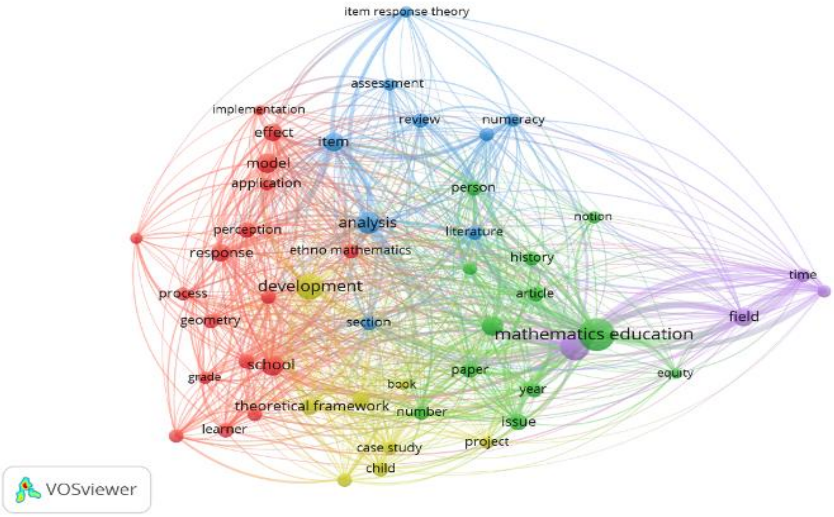


Fig 1. Network Visualization Vosviewers

In light of the background presented above, the objective of this study is to analyze DIF in an instrument designed to measure ethnomathematical knowledge among prospective mathematics teachers, with specific reference to the cultural dimensions of the Jambi region. To achieve this, the Rasch Model is implemented as the primary analytical framework, given its robustness in detecting item bias and ensuring psychometric fairness across culturally diverse groups.

2. RESEARCH METHOD

A. Research types and design

This study is a quantitative research (Soesana et al., 2023) that implements the Rasch Model to identify Differential Item Functioning (DIF). DIF occurs when items in a test behave differently for different groups, for example, based on culture or gender. Furthermore, this study uses an Ex Post Facto research design, where there is no manipulation of independent variables, but rather an analysis of data that has been collected. This model is suitable for identifying DIF patterns based on respondent groups such as cultural background or learning experience.

B. Rasch Model research model

In the DIF analysis using the Rasch Model in this study, the mathematical equation used is usually based on the one-parameter logistic model (1PL), which can be expressed as (Andrich, 2009):

$$P(X_{ni} = 1) = \frac{e^{(\theta n - \beta_i)}}{1+e^{(\theta n - \beta_i)}} \quad (1)$$

$P(X_{ni} = 1)$ is the probability that an individual nn answers item i correctly. θ_n is an individual's ability n. β_i is the difficulty level of the item i. In the context of DIF, differences in i item difficulty parameters between two groups (e.g., based on gender or educational background) may indicate bias in the questions. DIF in this study was further tested by comparing estimates of item difficulty parameters between different groups using the Likelihood Ratio Test (LRT) or Mantel-Haenszel Test.

Indonesian Journal of Education & Mathematical Science
Vol. 7, No. 2, Mei 2026, pp. 161~174
ISSN: 2715-985X, DOI: 10.30596/ijems.v7i2.29464

C. Research procedures

There are four main stages in this study, which include: Preliminary Research, Instrument Development and Permits, Data Collection and Analysis, Dissemination and Publication (Oriondo & Dallo-Antonio, 1984).

1) Preliminary research. This stage includes five main activities: first, analysis of mathematics learning, which identifies how ethnomathematics concepts have been integrated into the curriculum and teaching methods; second, analysis of Jambi culture, which traces cultural elements relevant to local-based mathematics learning; third, exploration of the Rasch Model concept, which includes parameter estimation mapping and instrument calibration mechanisms; fourth, DIF analysis, which focuses on detecting bias in test items based on respondent characteristics; and finally, analysis of the ethnomathematics concept, which links mathematical understanding with cultural perspectives to create more representative instruments. Through these stages, the study aims to develop a more reliable and fair measurement tool that can be used to improve the effectiveness of ethnomathematics teaching for prospective teachers in Jambi.
2) Instrument Development. At this stage, the research focused on developing a Rasch Model-based measurement tool to ensure validity and fairness in measuring the ethnomathematics knowledge of prospective mathematics teachers. This process began with the design of items based on ethnomathematics concepts and the cultural characteristics of Jambi, ensuring that the concepts were relevant to the learning context. Next, theoretical and empirical validity tests are conducted to assess the suitability of the instrument, followed by a preliminary analysis of the difficulties and abilities parameters on the Rasch scale to identify potential biases. The final development of the instrument includes determining the scoring rubric and DIF testing mechanism to ensure that the measuring tool provides fair results and can be used for more accurate knowledge evaluation. The results of this stage will form the basis for further analysis to improve the reliability and effectiveness of the evaluation of prospective mathematics teachers.
3) Collection and analysis of DIF data. At this stage, this study used the Rasch Model to identify DIF in the measurement instrument for prospective mathematics teachers' ethnomathematics knowledge. Data were collected through prospective teachers' responses to items designed based on the concepts of ethnomathematics and Jambi culture. After that, parameter estimation was carried out using the Rasch model to determine the levels of difficulty and ability, as well as to detect differences in item functions between respondent groups. DIF analysis was conducted by comparing parameter estimates between groups using Rasch analysis. The results of this analysis will form the basis for improving the instrument so that it is more optimal in measuring ethnomathematics knowledge validly and reliably. DIF data analysis in this study was conducted using WINSTEP 3.73 software (Linacre, 2011).
4) Dissemination and publication. At this stage, research results are communicated through various media to ensure their impact on the development of ethnomathematics learning. Key findings are presented in academic forums, seminars, and conferences to obtain feedback and increase the relevance of the research. In addition, publications are made in reputable scientific journals to support the development of reliable measurement instruments for prospective mathematics teachers.

Indonesian Journal of Education & Mathematical Science
Vol. 7, No. 2, Mei 2026, pp. 161~174
ISSN: 2715-985X, DOI: 10.30596/ijems.v7i2.29464

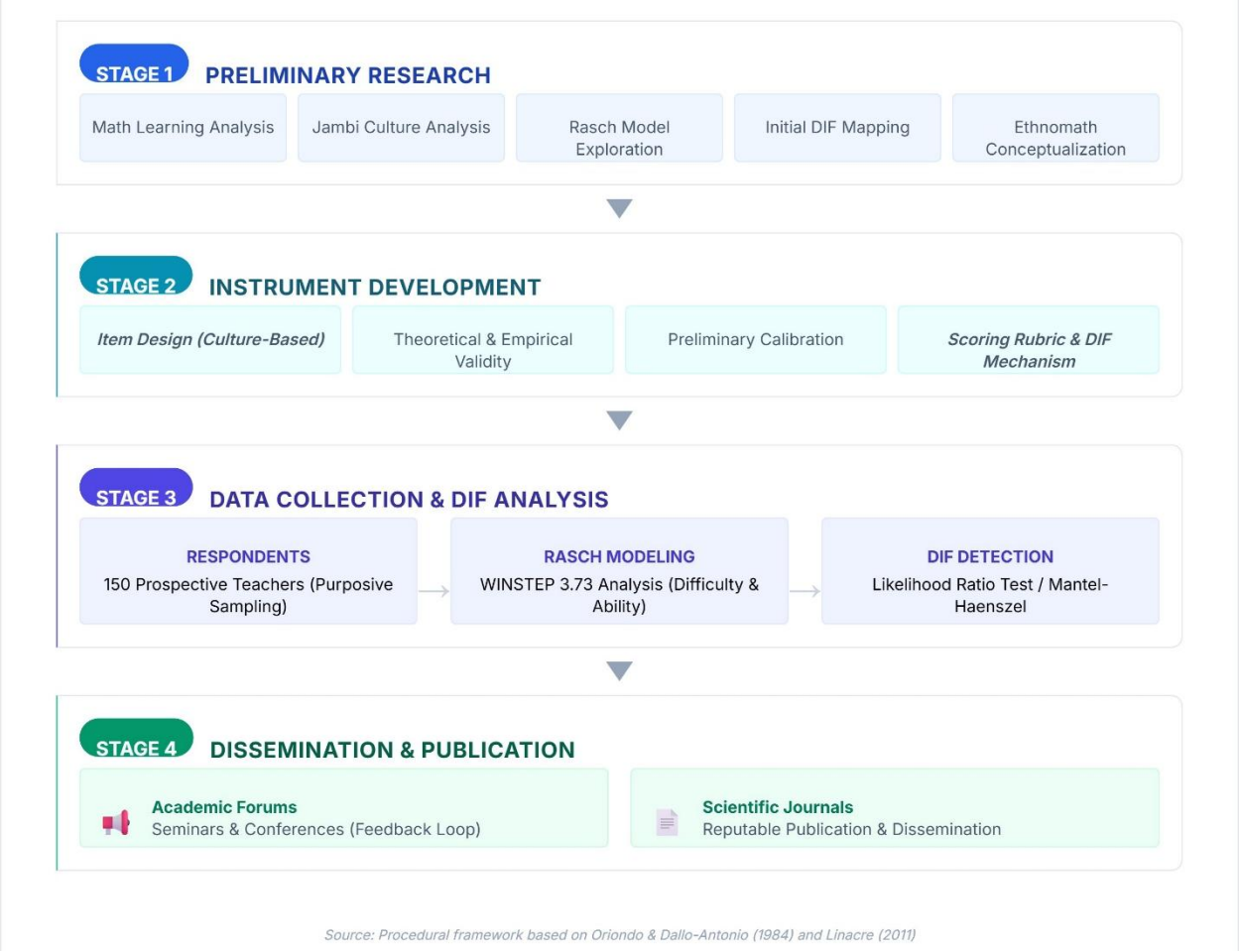


Fig 2. Research Procedural Flowchart: Etnomathematics DIF Analysis

D. Research respondents

The respondents or participants in this study were 150 prospective mathematics teachers, comprising mathematics education students and teacher education students from Jambi Province. The sampling technique employed was purposive sampling, ensuring that individuals selected had relevant academic backgrounds aligned with the study's objectives. Data collection was conducted using an ethnomathematics knowledge test instrument, which was designed to meet established criteria for validity and reliability.

3. RESULTS AND DISCUSSION

A. Descriptive statistics of respondents

This section presents descriptive statistics of the respondents, providing a comprehensive overview of their demographic and professional attributes. This section serves to contextualize the study by presenting key characteristics such as gender, educational attainment, and teaching experience within school environments. These variables are essential for understanding the composition of the sample and for interpreting subsequent findings with greater precision. By examining these foundational aspects, the study ensures analytical rigor and enhances the validity of its conclusions within the broader educational research framework.

Indonesian Journal of Education & Mathematical Science
Vol. 7, No. 2, Mei 2026, pp. 161~174
ISSN: 2715-985X, DOI: 10.30596/ijems.v7i2.29464

Table 1. Summary Table of Respondents' Descriptive Statistics

Category	Subcategory	Number of Respondents	Percentage (%)
Gender	Male	33	22%
	Female	117	78%
Educational Status	Undergraduate Student (S1)	71	48%
	Master's Student (S2)	31	21%
	Teacher Professional Education (PPG)	41	27%
	S1 Graduate	5	3%
	S2 Graduate	2	1%
Teaching Experience	Has taught in an educational institution	89	59%
	Has not taught in an educational institution	61	41%
Total Respondent		150	100%

The demographic analysis of the 150 respondents reveals a significant gender disparity, with female participants (78%) substantially outnumbering males (22%). In terms of educational attainment, the majority of the sample consists of undergraduate students (48%), followed by those enrolled in teacher professional education (27%) and master's programs (21%), while a small fraction (4%) had already completed their degrees. Furthermore, professional background data indicates that 59% of respondents possess prior teaching experience in formal educational institutions, whereas 41% have yet to engage in formal pedagogical roles.

The predominance of female respondents in this study is consistent with the global trend toward the "feminization of the teaching profession," particularly in mathematics and primary education. This finding aligns with the research by Nicol et al., (2024), which highlights that female preservice teachers often constitute the largest demographic in teacher education cohorts. From a psychometric standpoint, this imbalance underscores the need to conduct Differential Item Functioning (DIF) analysis to ensure that ethnomathematical assessment instruments do not harbor latent gender biases that could compromise the validity of competency measures. Furthermore, the substantial representation of participants with prior teaching experience (59%) provides a unique vantage point for assessing ethnomathematical knowledge. Unlike studies that rely solely on inexperienced undergraduates, this diverse sample allows for a more nuanced investigation into how professional experience mediates the interpretation of cultural mathematics. This result reinforces the arguments of Rosa et al., (2016), who posit that pedagogical readiness is a product of the intersection between academic theory and field-based practice. By including both novice and experienced practitioners, this study strengthens the generalizability of the Rasch-based findings, positioning the instrument as a reliable tool for diverse pedagogical contexts in the Indonesian educational landscape.

B. Summary of the measured item and person

This subsection provides a summary of the measured items and persons, offering a foundational overview of the data collected through the assessment instrument. The analysis includes key metrics

Indonesian Journal of Education & Mathematical Science
Vol. 7, No. 2, Mei 2026, pp. 161~174
ISSN: 2715-985X, DOI: 10.30596/ijems.v7i2.29464

that reflect both item-level performance and individual respondent outcomes, enabling a comprehensive evaluation of the instrument's psychometric properties.

Table 2 presents a comprehensive summary of the measurement results for 150 individuals, highlighting key statistical indicators relevant to the assessment's psychometric quality. The data include total scores, person measures, model error, and fit statistics (Infit and Outfit), both in terms of mean square (MNSQ) and standardized values (ZSTD). The mean total score was 31.8, with a standard deviation of 5.8, indicating a moderate spread of performance across respondents. The person measure ranged from -2.68 to 3.92 logits, suggesting a wide range of ability levels within the sample.

Table 2. Summary of 150 Measured Person

	Total Score	Count	Measure	Model Error	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	31.8	43.0	1.70	0.47	0.96	0.0	1.10	0.2
S.D	5.8	0.0	1.11	0.08	0.26	1.0	1.26	1.3
Max.	40.0	43.0	3.92	0.74	1.75	3.3	9.14	6.4
Min	6.0	43.0	-2.68	0.38	0.51	-2.3	0.14	-1.5

Person reliability = 0.81
Cronbach Alpha (KR-20) person Raw Score "Test" reliability = 0.85

The fit statistics further support the internal consistency of the measurement model. The mean Infit MNSQ was 0.96 and Outfit MNSQ was 1.10, both of which fall within the acceptable range of 0.5 to 1.5, indicating that the data fit the Rasch model reasonably well. The corresponding ZSTD values (0.0 for Infit and 0.2 for Outfit) reinforce this interpretation, as they remain close to the expected value of zero, suggesting minimal distortion in the measurement process. The psychometric findings of this study provide empirical weight to the argument that the Rasch model offers a more stable framework for ethnomathematical assessment compared to Classical Test Theory (CTT). The Infit and Outfit MNSQ values (0.96–1.10) obtained in this study are strikingly consistent with the research of Sumintono & Widhiarso (2014), who emphasize that fit indices within this range guarantee that the items are productive for measurement. This stability suggests that the ethnomathematical tasks designed for preservice teachers functioned as a unidimensional construct, effectively minimizing construct-irrelevant variance.

Moreover, the high item reliability (0.97) and person reliability (0.81) reinforce the findings of Herdoost (2016) and Irwing et al., (2018), which suggest that reliability coefficients above 0.80 indicate an instrument's readiness for high-stakes educational diagnostics. This high level of precision is critical when assessing "ethnomathematical knowledge," a domain often criticized for its subjective nature. By achieving a Cronbach's Alpha of 0.85, this study strengthens the position of Biddix (2017), proving that cultural-based mathematics items can achieve the same, if not higher, level of dependability as standardized curricula items.

Table 3 provides a detailed summary of 43 measured items, offering key insights into the psychometric properties of the instrument used in this study. The table includes descriptive statistics such as total score, item count, item measure, model error, and fit indices (Infit and Outfit), both in mean square (MNSQ) and standardized (ZSTD) formats. These metrics collectively serve to evaluate the quality and consistency of item performance across the sample of 150 respondents (Sumintono & Widhiarso, 2014).

The mean total score for the items was 111.0, with a standard deviation of 35.8, indicating a broad distribution of response values. All items were administered uniformly, as reflected by the consistent count of 150. A notable point of discussion is the wide logit range (-2.89 to 5.02), which demonstrates an excellent "item-person map" alignment. This spread reflects the principles discussed

by Fischer et al., (1999) and Van Der Linden (2010), where a balanced distribution of difficulty ensures the instrument's construct validity. Compared to previous studies that often showed "ceiling effects" (where items are too easy for preservice teachers), the current instrument successfully captures the entire spectrum of ethnomathematical proficiency. This alignment ensures that the subsequent Differential Item Functioning (DIF) analysis is performed on a statistically sound foundation, providing a credible benchmark for future ethnomathematical research in diverse cultural settings

Fit statistics further reinforce the integrity of the measurement model. The average Infit MNSQ was 0.99 and Outfit MNSQ was 1.10, both of which fall within the acceptable range of 0.5 to 1.5, indicating that the items functioned as expected under the Rasch model. The corresponding ZSTD values (0.0 for both Infit and Outfit) suggest minimal deviation from model expectations, with only minor fluctuations observed in the standard deviations.

Most notably, the item reliability coefficient was 0.97, which reflects an exceptionally high level of internal consistency. This value indicates that the items reliably differentiate among respondents and that the instrument is stable across repeated administrations. Such a high reliability index is critical for ensuring the validity of subsequent analyses and for supporting the generalizability of the findings. In sum, the data presented in Table 3 affirm the robustness of the measurement instrument and its suitability for rigorous empirical investigation (Taherdoost, 2016).

Table 3. Summary of 43 Measured Item

	Total Score	Count	Measure	Model Error	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	111.0	150.0	0.00	0.29	0.99	0.0	1.10	0.0
S.D	35.8	0.0	1.83	0.11	0.16	1.5	1.00	1.7
Max.	147.0	150.0	5.02	0.61	1.37	4.4	5.44	5.1
Min	8.0	150.0	-2.89	0.18	0.73	-2.7	0.19	-2.7

Item reliability = 0.97

Indonesian Journal of Education & Mathematical Science
Vol. 7, No. 2, Mei 2026, pp. 161~174
ISSN: 2715-985X, DOI: 10.30596/ijems.v7i2.29464

C. Differential item functioning classification

Table 4. DIF Summary Based on Gender, Education Level, and Teaching Experience

Item Number	Gender			Education level			Teaching experiences		
	DIF χ^2	Prob.	t=ZSTD	DIF χ^2	Prob.	t=ZSTD	DIF χ^2	Prob.	t=ZSTD
1	0,27	0,60	-0,66	1,94	0,38	-0,45	2,75	0,10	0,71
2	1,79	0,18	0,17	2,63	0,45	-0,92	1,16	0,28	0,10
3	0,05	0,82	-1,08	0,82	0,84	-1,73	0,01	0,92	-1,25
4	0,65	0,42	-0,29	2,31	0,31	-1,26	0,00	1,00	-1,47
5	5,45	0,02	1,07	7,91	0,09	-0,42	0,00	1,00	-1,53
6	0,42	0,52	-0,54	8,77	0,07	-0,17	2,31	0,13	0,54
7	2,71	0,10	0,59	3,41	0,18	-0,37	0,00	1,00	-1,46
8	4,84	0,03	1,01	1,12	0,57	-0,85	1,45	0,23	0,27
9	1,36	0,24	0,07	1,23	0,54	-0,91	0,07	0,80	-0,97
10	3,56	0,06	0,64	1,95	0,58	-1,53	0,08	0,78	-0,94
11	0,67	0,41	-0,34	2,77	0,43	-1,32	0,04	0,85	-1,10
12	0,01	0,92	-1,29	0,58	0,45	-0,45	0,72	0,40	-0,13
13	0,26	0,61	-0,70	5,50	0,24	-0,68	2,10	0,15	0,49
14	0,59	0,44	-0,37	1,22	0,75	-2,05	0,24	0,63	-0,61
15	7,10	0,01	1,33	1,98	0,74	-2,36	0,22	0,64	-0,65
16	0,34	0,56	-0,56	1,31	0,52	-1,04	0,09	0,77	-0,90
17	0,00	1,00	0,00	0,37	0,54	-0,26	1,57	0,21	0,33
18	0,00	1,00	0,00	1,91	0,38	-0,60	0,95	0,33	0,02
19	3,06	0,08	0,63	5,53	0,06	0,22	0,02	0,90	-1,22
20	0,77	0,38	-0,23	4,14	0,12	-0,17	0,09	0,77	-0,90
21	0,03	0,86	-1,13	1,82	0,77	-2,57	1,51	0,22	0,26
22	0,02	0,90	-1,25	2,03	0,36	-0,49	0,87	0,35	-0,04
23	1,22	0,27	-0,04	8,01	0,05	0,21	3,10	0,08	0,80
24	0,75	0,39	-0,26	1,99	0,37	-0,64	0,19	0,67	-0,69
25	0,21	0,65	-0,77	0,96	0,92	-2,45	0,25	0,62	-0,59
26	1,66	0,20	0,22	1,34	0,25	0,46	2,12	0,15	0,55
27	0,75	0,39	-0,28	3,67	0,30	-0,82	0,30	0,59	-0,53
28	5,89	0,02	1,19	2,52	0,28	-0,29	0,00	0,95	-1,37
29	2,11	0,15	0,31	5,26	0,15	-0,60	2,97	0,08	0,78
30	0,00	1,00	0,00	0,34	0,56	-0,67	0,15	0,70	-0,75
31	0,66	0,42	-0,31	3,15	0,20	-0,09	0,15	0,70	-0,76
32	0,44	0,51	-0,50	2,61	0,46	-1,28	3,79	0,05	1,00
33	0,50	0,48	-0,46	1,15	0,77	-1,80	1,32	0,25	0,19
34	0,38	0,54	-0,60	1,69	0,64	-1,63	0,18	0,67	-0,71
35	0,68	0,41	-0,33	1,08	0,03	0,16	0,42	0,52	-0,40
36	3,19	0,07	0,69	0,00	1,00	0,00	1,12	0,29	0,11
37	0,89	0,35	-0,21	6,79	0,08	0,00	0,82	0,37	-0,09
38	0,37	0,54	-0,53	0,00	1,00	0,00	0,25	0,62	-0,59
39	1,71	0,19	0,22	0,85	0,65	-1,26	0,12	0,73	-0,81
40	0,25	0,62	-0,74	0,97	0,61	-0,94	2,54	0,11	0,62
41	0,65	0,42	-0,29	0,61	0,43	0,01	2,62	0,11	0,73
42	0,07	0,78	-0,98	0,00	1,00	0,00	0,00	1,00	0,00

Indonesian Journal of Education & Mathematical Science
Vol. 7, No. 2, Mei 2026, pp. 161~174
ISSN: 2715-985X, DOI: 10.30596/ijems.v7i2.29464

43	0,00	1,00	-1,44	0,97	0,61	-1,16	0,00	1,00	-1,51

Table 4 presents a DIF analysis based on three respondent characteristics: gender, education level, and teaching experience. The table reports DIF statistics for each item, including the Chi-Square value, associated probability (Prob.), and standardized t-value (ZSTD). These indicators are critical for evaluating whether items function equivalently across subgroups, thereby ensuring fairness and validity in the measurement instrument (Naga, 2012)

The probability values (Prob.) serve as the primary reference for determining statistical significance in DIF analysis. A conventional threshold of $p < 0.05$ is typically used to identify items with significant DIF. Based on this criterion, several items exhibit notable DIF across gender, such as Item 5 (Prob. = 0.02), Item 8 (Prob. = 0.03), Item 15 (Prob. = 0.01), and Item 28 (Prob. = 0.02). These findings suggest that responses to these items may be influenced by gender-related factors, warranting further investigation into potential bias or differential interpretation.

In contrast, DIF based on education level and teaching experience appears less pronounced. Although some items show relatively low probability values, such as Item 23 (Prob. = 0.05) and Item 32 (Prob. = 0.05) for teaching experience, most items yield probabilities well above the significance threshold, indicating minimal DIF. For education level, while Item 35 shows a very low probability (Prob. = 0.03), the corresponding ZSTD value (0.16) remains within acceptable bounds, suggesting limited practical impact.

The ZSTD values complement the probability analysis by indicating the magnitude and direction of DIF. Values exceeding ±2.0 may signal substantial DIF, yet most items in this table fall within the range of −1.5 to +1.5, reinforcing the overall stability of item functioning across subgroups. Taken together, the DIF analysis supports the general fairness of the instrument, with only a few items requiring closer scrutiny to ensure measurement equivalence across gender, education level, and teaching experience. This reinforces the psychometric integrity of the instrument and its suitability for use in diverse respondent populations (Rust & Golombok, 2018).

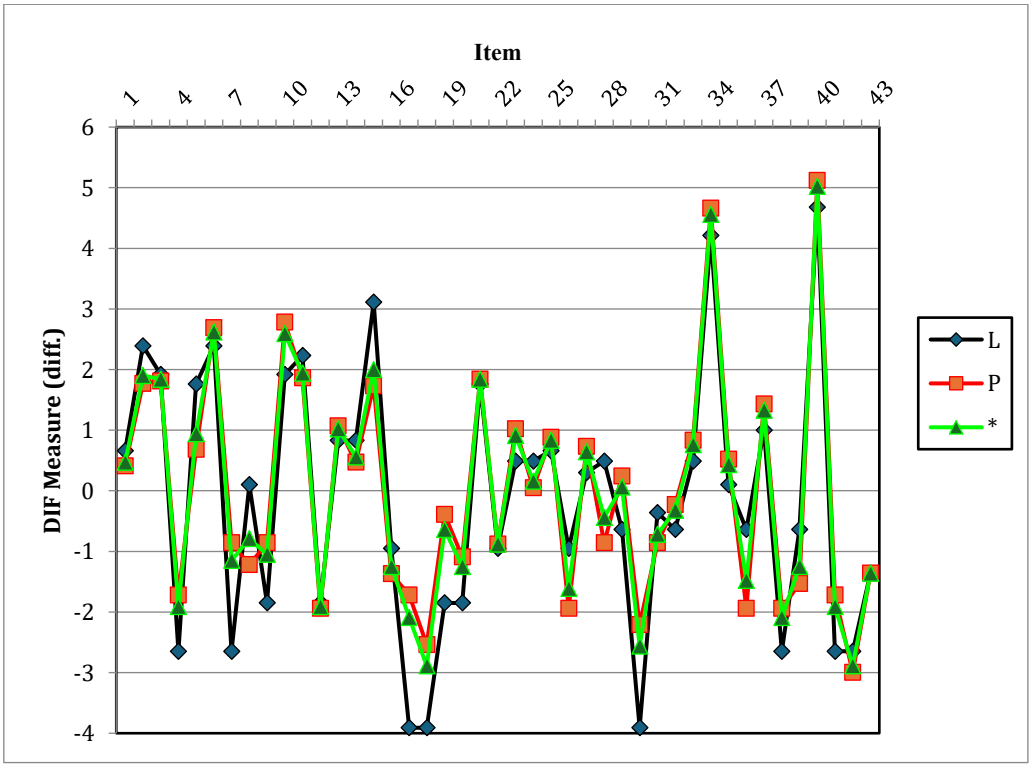


Fig 3. Person DIF Plot Based on Gender.

Indonesian Journal of Education & Mathematical Science
Vol. 7, No. 2, Mei 2026, pp. 161~174
ISSN: 2715-985X, DOI: 10.30596/ijems.v7i2.29464

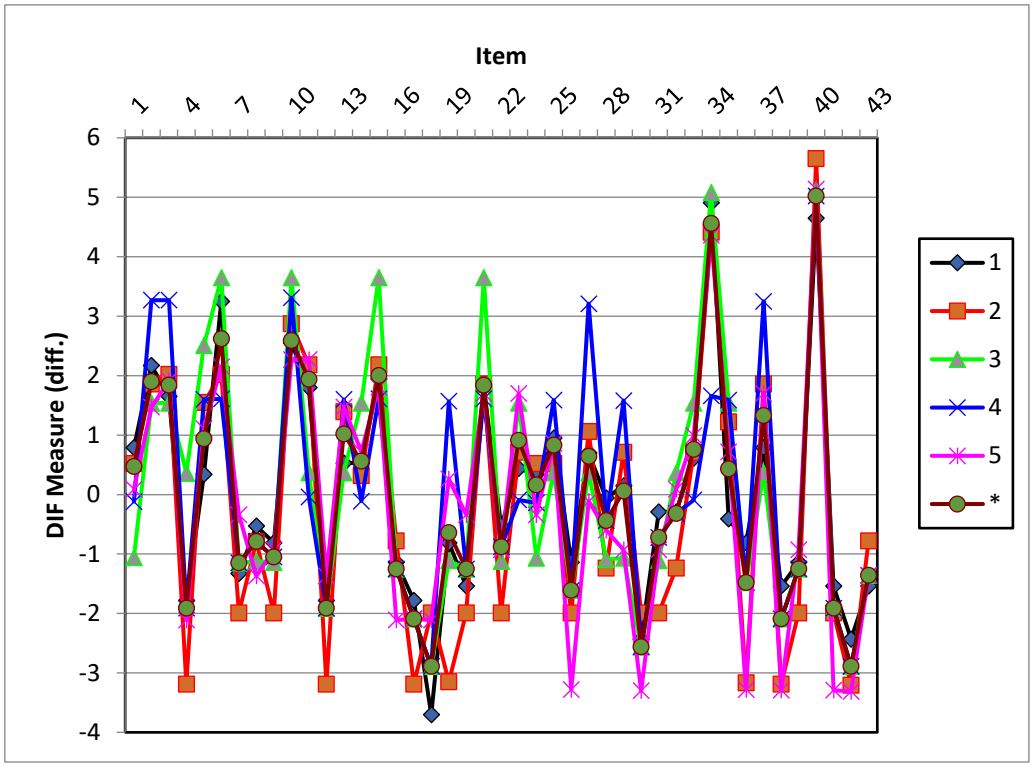


Fig 4. Person DIF Plot Based on Teaching Experiences

Figures 3 and 4 provide a detailed graphical representation of DIF across 43 items, facilitating a nuanced examination of measurement invariance at the item level among gender groups and multiple respondent categories. These visualizations serve as essential psychometric tools in evaluating the fairness and validity of assessment instruments, particularly in contexts where construct comparability across subgroups is critical (Kline, 2015).

Figure 3 depicts DIF measures by gender, with three distinct curves represented by blue diamonds (L), red squares (P), and green triangles. The green curve reveals substantial variability, with several items exceeding the ±3 threshold, indicating potential differential functioning across gender groups. Such deviations may reflect systematic bias or differences in item interpretation influenced by gender-specific experiences. In contrast, the blue and red curves exhibit relative stability, suggesting consistent item performance and supporting the assumption of measurement equivalence for those subgroups. Items associated with extreme DIF values in the green curve merit further scrutiny to ensure content neutrality and linguistic clarity (Cohen et al., 1993; Rustam et al., 2019).

Figure 4 extends the analysis to six respondent groups, each distinguished by unique symbols and colors. The DIF values range from –5 to +5, offering a broader scale for detecting item sensitivity. Several curves intersect and diverge markedly, suggesting that certain items may not function uniformly across groups. This pattern points to potential construct-irrelevant variance, which could undermine the validity of score interpretations. Items consistently demonstrating high DIF across multiple groups should be prioritized for revision, as they may be influenced by latent cultural, contextual, or cognitive factors (Dorans et al., 2013).

Indonesian Journal of Education & Mathematical Science
Vol. 7, No. 2, Mei 2026, pp. 161~174
ISSN: 2715-985X, DOI: 10.30596/ijems.v7i2.29464

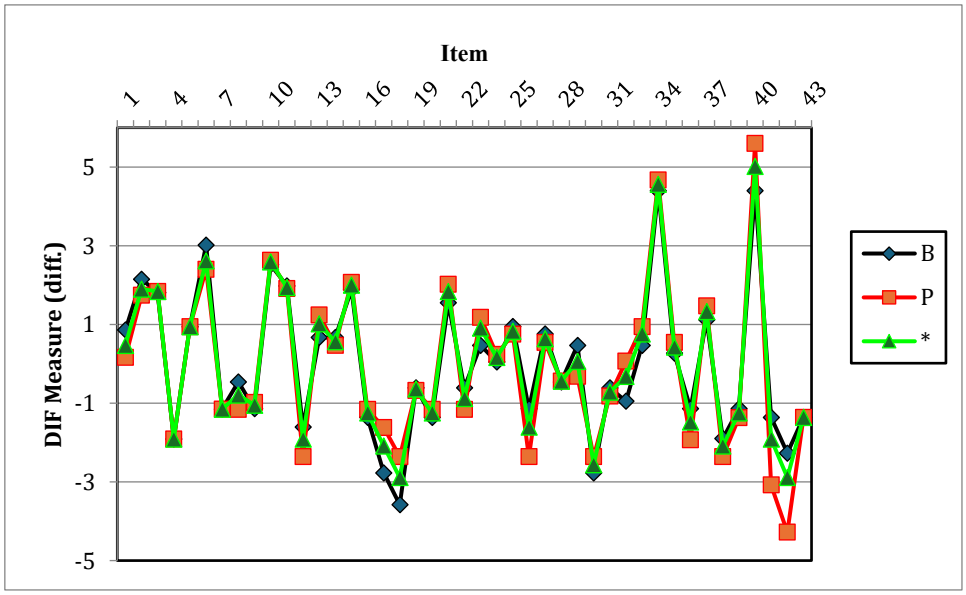


Fig. 5. Person DIF Plot Based on Education Level

Figure 5 refines the DIF analysis by comparing three additional groups: blue diamonds (B), red squares (P), and green triangles. Once again, the green curve exhibits the most fluctuation, with a prominent peak at item 37 and a sharp decline at item 41. These variations suggest that specific items may be interpreted differently depending on group affiliation, potentially affecting the fairness of the measurement. The relative consistency of the blue and red curves reinforces the reliability of those items within their respective groups, while the volatility observed in the green curve underscores the need for targeted psychometric evaluation.

Taken together, the three DIF plots highlight the importance of rigorous item-level analysis in ensuring equitable measurement across diverse respondent populations. Items flagged for extreme DIF should undergo systematic content review and, where necessary, be revised to enhance interpretive consistency. This approach is particularly relevant in the context of assessing preservice teachers' ethnomathematical knowledge, where cultural and contextual factors may influence item comprehension. The findings contribute to the refinement of Rasch-based instruments and support the development of valid, bias-free tools for evaluating domain-specific competencies in teacher education.

Despite the methodological rigor and psychometric robustness demonstrated in this study, several limitations warrant acknowledgment. First, the use of purposive sampling, while appropriate for targeting preservice mathematics teachers familiar with ethnomathematical content, may restrict the generalizability of findings beyond the Jambi region. Second, the reliance on quantitative DIF analysis, though effective in detecting statistical item bias, does not capture the nuanced cognitive and cultural interpretations that may underlie DIF. Third, the absence of qualitative triangulation, such as think-aloud protocols or expert validation panels, limits the depth of insight into item-level discrepancies. Additionally, the study's focus on three respondent characteristics (gender, education level, and teaching experience) may overlook other influential variables, such as linguistic background or regional dialects, which could affect item comprehension. Future research should address these limitations by employing mixed-method approaches, expanding the respondent base across diverse cultural settings, and incorporating qualitative techniques to enrich the interpretive validity of DIF findings.

4. CONCLUSION

In conclusion, this research transcends a mere summary of psychometric outputs by elucidating the critical tension between cultural contextualization and measurement invariance in the measurement of ethnomathematical knowledge among preservice teachers. While the Rasch Model successfully validated the instrument's overall structure, the identification of significant Differential Item Functioning (DIF) provides a pivotal theoretical contribution: it challenges the monolithic assumption that ethnomathematical content inherently ensures assessment equity. These findings suggest that

cultural stimuli are not interpreted uniformly; rather, they are filtered through diverse socio-cognitive lenses shaped by gender, educational background, and prior experience, which can inadvertently introduce construct-irrelevant variance even within culturally grounded tasks. Practically, this underscores a necessary paradigm shift in the development of culture-based assessments, where "contextual relevance" must be strictly balanced with "statistical neutrality" through rigorous DIF screening to prevent localized biases from undermining teacher competency evaluations. By establishing a "bias-aware" framework for ethnomathematical measures, this study offers a robust blueprint for teacher education programs to develop equitable instruments that are both mathematically rigorous and culturally sensitive. Ultimately, while this study provides a foundational step in ensuring fairness, the transition from detecting to explaining these biases necessitates future qualitative integration, such as cognitive interviews, to ensure that ethnomathematical assessments truly reflect mathematical proficiency rather than demographic affiliation across the heterogeneous landscape of mathematics education.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Universitas Jambi for funding this research.

REFERENCES

Alkharusi, H. (2020). Development and Psychometric Properties of a Mathematical Literacy Scale. *International Journal of Instruction*, *13*(3), 621–636. https://doi.org/10.29333/iji.2020.13342a

Andrich, D. (2009). *Rasch Model for Measurement*. Sage University.

Atmaja, T. S. (2023). The Urgency Of Character Education In Educational Units In Indonesia In Facing Global Challenges And 21st Century Competencies. *Jurnal Scientia*, *12*(04), 2014–2019.

Bahar, M., Wahyuni, A., & Saputra, M. A. (2023). Educational Values in Lubuk Larangan Local Wisdom in Rantau Pandan Village, Bungo Regency, Jambi Province. *QALAMUNA: Jurnal Pendidikan, Sosial, Dan Agama*, *15*(2), 999–1006.

Biddix, P. (2017). Instrument, Validity, Reliability | Research Rundowns. In *University of Missouri - St. Louis* (p. 3). https://researchrundowns.com/quantitative-methods/instrument-validity-reliability/

Cohen, A. S., Kim, S., & Baker, F. B. (1993). Detection of Differential Item Functioning in the Graded Response Model. *Applied Psychological Measurement*, *17*(4), 335–350.

Desai, S., Kurtz, B., & Safi, F. (2021). Mathematics Heritage Project: An Exploration Empowering Students' Mathematical Identities. *Journal of Humanistic Mathematics*, *11*(2), 106–122. https://doi.org/10.5642/jhummath.202102.05

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. In *Journal of Business Research* (Vol. 133, pp. 285–296). https://doi.org/10.1016/j.jbusres.2021.04.070

Dorans, N. J., Carlson, J., Bridgeman, B., Golub-smith, M., & Greenwood, R. (2013). *Constructed-Response DIF Evaluations for Mixed-Format Tests*. *December*.

Ergene, Ö., Ergene, B. Ç., & YAZICI, E. (2020). Ethnomathematics activities: reflections from the design and implementation process. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *11*(2), 402–437.

Falani, I., Iriyadi, D., Wardani, I. Y., Susanti, H., & Nasution, R. A. (2022). A Rasch Analysis of Perceived Stigma of Covid-19 among Nurses in Indonesia Questionnaire. *Psychological Thought*, *15*(1), 12–28. https://doi.org/10.37708/psyct.v15i1.530

Fischer, G. H., Molenaar, I. W., & Ivo W, M. (1999). *Rasch Models Foundations, Recent Developments, and Applications*. Springer. https://doi.org/10.1007/978-1-4612-4230-7

Herdoost, H. (2016). Validity and Reliability of the Research Instrument ; How to Test the Validation of a Questionnaire / Survey in a Research Hamed Taherdoost To cite this version : HAL Id : hal-02546799 Validity and Reliability of the Research Instrument ; How to Test the. *International Journal of Academic Research in Management*, *5*(3), 28–36.

Irwing, P., Booth, T., & Hughes, D. J. (2018). *The Wiley Handbook of Psychometric Testing*. John Wiley & Sons Ltd.

Keliat, B. A., Nasution, R. A., Handini, I. T., & Falani, I. (2023). Psychometric properties evaluation of the Indonesian version of the Beck Scale for Suicide Ideation (BSSI) questionnaire using a Rasch model. *European Review for Medical & Pharmacological Sciences*, *27*(21).

Kline, P. (2015). *A Handbook of Test Construction: Introduction to psychometric design*.

Linacre, J. M. (2011). *Winsteps Help for Rasch Analysis*. http://homes.jcu.edu.au/~edtgb/%5Cnpapers3://publication/uuid/D56B724A-62FF-4D00-84E1-ECC888298B70

Mania, S., & Alam, S. (2021). Teachers' perception toward the use of ethnomathematics approach in teaching math. *International Journal of Education in Mathematics, Science and Technology*, *9*(2), 282–298.

Masters, G. N. (2020). Measuring Student Growth and Progress in the 21st Century. *Educational Measurement: Issues and Practice*, *39*(3), 56–62. https://doi.org/10.1111/emip.12358

Mosimege, M., & Egara, F. (2023). Students' perceptions of mathematics teachers' usage of ethnomathematics materials in the teaching and learning of mathematics. *International Journal of Multicultural Education*, *25*(2), 443–459.

Naga, D. S. (2012). *Teori Sekor pada Pengukuran Mental*. PT Nagarani Citrayasa.

Nicol, C., (Gels), C. N., & Peng, A. (2024). *Ethnomathematics and Mathematics Education*. Springer.

Nur, A. S., Waluya, S. B., Rochmad, R., & Wardono, W. (2020). Contextual Learning with Ethnomathematics in Enhancing the Problem Solving Based on Thinking Levels. *Journal of Research and Advances in Mathematics Education*, *5*(3),

331–344.
Oriondo, L. L., & Dallo-Antonio, E. M. (1984). *Evaluating educational outcomes: Tests, measurement and evaluation*. Rex Book Store.
Prahmana, R. C. I. (2022). Ethno-realistic mathematics education: The promising learning approach in the city of culture. *SN Social Sciences*, *2*(12), 257.
Putra, E. C. S., & Mahmudah, F. N. (2021). The implementation of ethnomathematics based-learning for students. *SJME (Supremum Journal of Mathematics Education)*, *5*(2).
Ramalisa, Y., Falani, I., & Pasaribu, F. T. (2023). Rasch analysis in developing Jambi culture-based ethnomathematics test for prospective mathematics teachers. *JRAMathEdu (Journal of Research and Advances in Mathematics Education)*.
Retnawati, H., Hadi, S., & Nugraha, A. C. (2018). Vocational High School Teachers' Difficulties in Implementing the Assessment in Curriculum 2013 in Indonesia. *International Journal of Instruction*, *11*(1), 33–48.
Rosa, M., & Orey, D. C. (2021). An ethnomathematical perspective of stem education in a glocalized world. *Bolema - Mathematics Education Bulletin*, *35*(70), 840–876. https://doi.org/10.1590/1980-4415v35n70a14
Rosa, M., Shirley, L., Gavarrete, M. E., & Alangui, W. V. (2016). *Ethnomathematics and its diverse approaches for mathematics education*. Springer.
Rust, J., & Golombok, S. (2018). Modern Psychometrics. In *Modern Psychometrics*. https://doi.org/10.4324/9781315787848
Rustam, A., Santun Naga, D., & Supriyati, Y. (2019). Sensitivity and Accuracy of the Mantel-Haenszel Method and Standardization Method: Detection of Item Functioning Differential. *International Journal of Education and Literacy Studies*, *7*(3), 28. https://doi.org/10.7575/aiac.ijels.v.7n.3p.28
Salam, M., Wahyuni, A., & Hajri, P. (2022). The Traditional Values of Maanta in Jambi Seberang Community. Al. *Hayat*, *6*(2).
Skaggs, G., & Williams, T. (2023). *Comparison of The Item Response Theory with Covariates Model and Explanatory Cognitive Diagnostic Model for Detecting and Explaining Differential Item Functioning*.
Soesana, A., Subakti, H., Karwanto, K., Fitri, A., Kuswandi, S., Sastri, L., Falani, I., Aswan, N., Hasibuan, F. A., & Lestari, H. (2023). *Metodologi Penelitian Kuantitatif*.
Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Trim Komunikata Publishing House.
Sunzuma, G., & Maharaj, A. (2019). Teacher-related challenges affecting the integration of ethnomathematics approaches into the teaching of geometry. *Eurasia Journal of Mathematics, Science and Technology Education*, *15*(9). https://doi.org/10.29333/ejmste/108457
Taherdoost, H. (2016). Validity and Reliability of the Research Instrument. In *SSRN Electronic Journal* (Vol. 5, Issue 3, pp. 28–36).
Templin, J. (n.d.). *Differential Item Functioning*.
Van Der Linden, W. J. (2010). Item response theory. *International Encyclopedia of Education*, *4*, 81–88. https://doi.org/10.1016/B978-0-08-044894-7.00250-5