REID (Research and Evaluation in Education), 11(2), 2025, 129-141
Available online at: http://journal.uny.ac.id/index.php/reid

Assessment and measurement bias in madrasa performance evaluation: Evidence from underdeveloped areas in Indonesia

Eko Wahyunanto Prihono*[1]; Ridhwan Latuapo[1]; Fitria Lapele[1]; Siti Irene Astuti Dwiningrum[2]; Janu Arlinwibowo[3]; Margarito Surbito Reyes Jr.[4]
[1]Universitas Islam Negeri Abdul Muthalib Sangadji Ambon, Indonesia
[2]Universitas Negeri Yogyakarta, Indonesia
[3]Badan Riset dan Inovasi Nasional, Indonesia
[4]Basilan National High School, Philippines
*Corresponding Author. E-mail: ekoprihono@iainambon.ac.id

ARTICLE INFO

Article History
Submitted:
February 20, 2024
Revised:
June 21, 2025
Accepted:
December 8, 2025

Keywords
assessment; inequality; madrasa performance; underdeveloped areas

Scan Me:

ABSTRACT

Inequality is a condition characterized by an unbalanced assessment process. Physical and psychological factors, both those measuring and those being measured, may impact assessment inequality. The purpose of this research was to highlight the potential inequity in the performance assessment of madrasas in underdeveloped areas. A quantitative research design was employed. The data were collected using a questionnaire instrument that had been proven valid and reliable. Path analysis was used to determine both direct and indirect effects. The findings showed that measurement errors related to the instruments used have a direct positive effect on inequality in the performance assessment of madrasas in underdeveloped areas, as well as an indirect effect mediated through teacher quality. One alternative solution to reducing the imbalance in assessing the performance of madrasas in underdeveloped areas can be implemented through policy dimensions, including macro, meso, and micro dimensions.

To cite this article (in APA style):
Prihono, E. W., Latuapo, R., Lapele, F., Dwiningrum, S. I. A., Arlinwibowo, J., & Reyes Jr., M. S. (2025). Assessment and measurement bias in madrasa performance evaluation: Evidence from underdeveloped areas in Indonesia. *REID (Research and Evaluation in Education)*, *11*(2), 129-141. https://doi.org/10.21831/reid.v11i2.71444

INTRODUCTION

Educational inequality in Indonesia is a multidimensional issue influenced by several factors, including the quality of human resources, educational opportunities, access, infrastructure, literacy, and learning processes (Anita & Dwiningrum, 2022). These disparities are further widened by regional development imbalances, particularly between urban and rural areas (Dwiningrum, 2007; Khilji et al., 2022). In the digital era, inequality also manifests through unequal access to digital information, with urban internet access surpassing 50% in 2019, compared to only about 30% in rural areas (Bida, 2021; Prihono, Retnawati, et al., 2022). This digital divide impacts access to educational resources and the success of digital learning initiatives promoted by the Ministry of Education and Culture (Hermawansyah, 2021). For underdeveloped regions, such conditions exacerbate challenges in delivering quality education (Prihono et al., 2025).

In the Indonesian education system, madrasas under the Ministry of Religious Affairs are formally equal to general schools (Wardi et al., 2019). Despite using the national curriculum, disparities persist between madrasas in developed versus underdeveloped areas. These include

limited infrastructure, low teacher quality and welfare, low student performance, and reduced relevance to both religious and general education (Bakhri, 2015; Huda, 2016). These dimensions reflect the core components used in evaluating madrasa performance: graduate quality, learning process, teacher quality, management, curriculum, assessment, infrastructure, and funding (N. Hasanah et al., 2025; Prihono et al., 2023; Wakano et al., 2024).

Madrasa performance is assessed through standardised accreditation instruments across all regions, including underdeveloped areas (E. Hasanah, 2021). However, uniform assessment tools may not accommodate contextual differences, potentially disadvantaging madrasas in underdeveloped areas. Based on Presidential Regulation No. 63 of 2020, Maluku Province includes six underdeveloped regencies: Maluku Tenggara Barat, Kepulauan Aru, Seram Bagian Barat, Seram Bagian Timur, Maluku Barat Daya, and Buru Selatan.

This inequality can result in the low performance value of madrasas in underdeveloped areas. Currently, there are 62 underdeveloped areas in Indonesia based on the Presidential Regulation No. 63 of 2020. One of the provinces that still has underdeveloped areas, based on the Presidential Decree, is Maluku Province. Maluku Province is recorded as having six underdeveloped regencies, including Maluku Tenggara Barat, Kepulauan Aru, Seram Bagian Barat, Seram Bagian Timur, Maluku Barat Daya, and Buru Selatan (Presidential Regulation No. 63 of 2020). Most of the results of evaluating the performance of madrasas in Maluku Province in 2021 are still low. This can be seen from the achievement of accreditation scores, where most of the madrasas obtained accreditation grades of C, and there were even non-accredited (NA or TT) madrasas. Details on the achievement of the madrasa performance assessment in terms of Maluku Province accreditation in 2021 are presented in Figure 1.

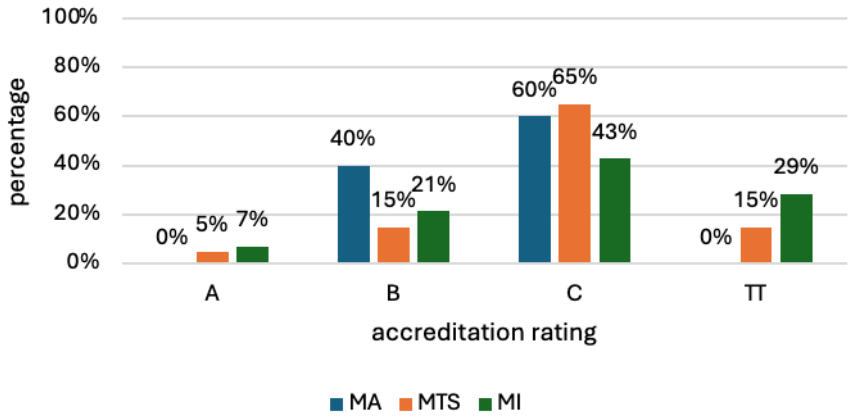

Figure 1. Results of Maluku Province Madrasa Performance Achievements

Based on Figure 1, data from 2021 show that many madrasas in Maluku Province received C accreditation or remained unaccredited. Specifically, performance data show that only 5% of madrasas achieved an A accreditation, 21% B, 56% C, and 18% were not accredited. At the MI level, 29% were not accredited. This suggests that 74% of madrasas are categorised as having low accreditation performance, reflecting deep-rooted structural inequalities. As defined by Jumono et al. (2021) and Sari et al. (2020), underdeveloped regions suffer from deficiencies in human capital, economy, and infrastructure.

While previous studies have addressed the challenges faced by madrasas in disadvantaged areas (Bakhri, 2015; Huda, 2016), they have not sufficiently explored how bias in assessment instruments or assessor subjectivity might contribute to inequality. Several studies (Anita & Dwiningrum, 2022; E. Hasanah, 2021) have recognised systemic disparities, but a gap remains in understanding the mechanisms by which assessment error (by assessors) and measurement error (by instruments) shape madrasa performance outcomes. In addition, there is limited research that applies statistical modelling, such as path analysis, to examine the direct and indirect effects of

these disparities on performance outcomes. The lack of empirical evidence addressing these latent error factors constitutes a significant gap that this study seeks to fill.

Theoretically, this study contributes to the discourse on educational equity by identifying latent factors contributing to accreditation-based performance disparities in madrasas. Practically, it offers policymakers insights into revising assessment instruments or implementing differentiated accreditation frameworks tailored to regional contexts. By integrating empirical models, this research provides an evidence-based foundation for educational reforms in underdeveloped regions.

Therefore, this study aims to highlight the potential inequities in the performance assessment of madrasas in underdeveloped areas of Maluku Province, using path analysis. Based on the conceptual framework and previous literature, the hypotheses tested in this study are:

H_1 : Assessment errors have a positive and significant effect on madrasa performance.
H_2 : Measurement error has a positive and significant effect on madrasa performance.
H_3 : Assessment errors have a positive and significant indirect effect on madrasa performance mediated by teacher quality.
H_4 : Measurement error has a positive and significant indirect effect on madrasa performance mediated by teacher quality.

METHOD

Research Design

This study employed a quantitative research design with a causal-comparative correlational approach to investigate the factors influencing madrasa performance in underdeveloped areas. The purpose of this approach was to examine both direct and indirect relationships among variables. Specifically, the study sought to investigate how assessment error and measurement error impact madrasa performance, both directly and through the mediating role of teacher quality. This design enabled the identification of patterns of inequality that may arise in the accreditation process, particularly in geographically and socioeconomically disadvantaged regions.

The study focused on four core variables, categorised as follows: (1) Assessment Error (AE) and (2) Measurement Error (ME) were treated as independent variables, meaning they were assumed to influence other variables but were not influenced by them within the model. (3) Teacher Quality (TQ) was treated as a mediating variable, while (4) Madrasa Performance (MP) was the ultimate dependent variable. The relationships among these variables were formulated into four hypotheses, which were statistically tested using path analysis. To visually represent the structure of these hypothesised relationships, the conceptual framework of the research is presented in Figure 2. This model was developed based on the literature review and illustrates both the direct and mediated paths among the key variables studied.

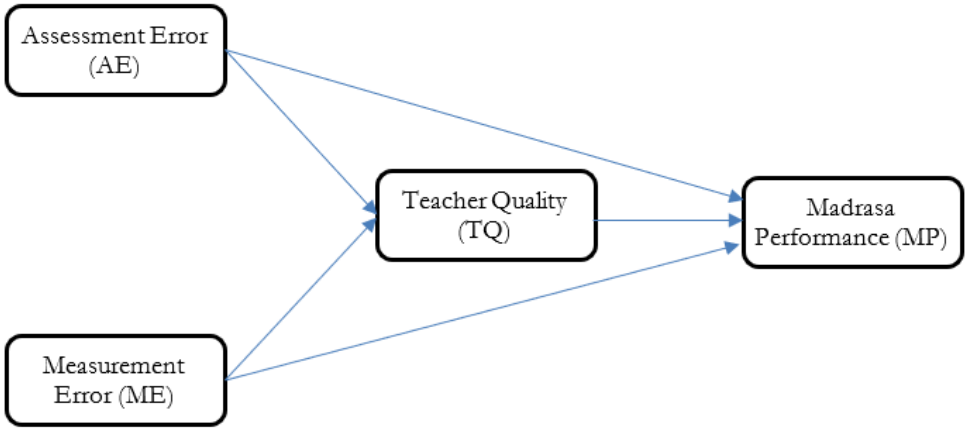

Figure 2. Research Conceptual Framework

The research was conducted in Maluku Province, an area that includes several districts designated as underdeveloped, based on the Presidential Regulation No. 63 of 2020. The study population consisted of madrasas that participated in the 2021 national accreditation process. A total of 18 madrasas across the education levels of *Madrasah Ibtidaiyah* (MI, primary education level), *Madrasah Tsanawiyah* (MTs, junior secondary education level), and *Madrasah Aliyah* (MA, senior secondary education level), were selected using purposive sampling. The sampling criteria ensured that each madrasa included had undergone a formal accreditation visitation during the specified year.

Each madrasa contributed five respondents, including one madrasa head, three teachers, and one school operator, resulting in a total sample of 90 individuals. These respondents were selected due to their direct involvement in the accreditation assessment process, ensuring that their perspectives and data were contextually grounded. In terms of role distribution, the sample comprised 20% *madrasa* heads (18 people), 60% teachers (54 people), and 20% school operators (18 people). The respondents' educational backgrounds ranged from undergraduate to post-graduate, reflecting diverse qualifications relevant to the study. Respondent characteristic data includes gender, age, and recent education. Details of the sample in this study are presented in Table 1.

Table 1. Research Samples

Category	Headmaster		Teacher		Operator		Whole	
	f	%	f	%	f	%	f	%
Gender								
Man	12	66.7	32	59.3	11	61.1	55	61.1
Woman	6	33.3	22	40.7	7	38.9	35	38.9
Age								
25-36 years	-	0	16	29.6	8	44.4	24	26.7
37-48 years	2	11.1	24	44.4	8	44.4	34	37.8
49-60 years	16	88.9	14	25.9	2	11.1	32	35.6
Educational Stage								
Diploma	-	0	2	3.7	9	50	11	12.2
Bachelor	6	33.3	48	88.9	8	44.4	62	68.9
Masters	12	66.7	4	7.41	1	5.56	17	18.9

The respondents in this study consisted of madrasa heads, teachers, and school operators directly involved in the accreditation process. A total of 90 respondents were selected purposively from 18 *madrasas* across different education levels (MI, MTs, and MA). Based on gender distribution, the majority of respondents were male (61.1%), suggesting that teaching positions in madrasas in underdeveloped areas tend to be occupied more by men.

In terms of age, respondents aged 25–36 years accounted for the lowest proportion (26.67%). This finding raises concerns about the potential shortage of teaching staff in the next five to 11 years, as older teachers retire without adequate generational replacement. Regarding educational qualifications, 12.22% of respondents held a diploma degree. These findings underscore systemic challenges in recruiting and retaining qualified educators in underdeveloped regions.

Research Instruments

The research instrument in this study was developed using a 4-point Likert scale, with response options ranging from 1 (Not Relevant) to 4 (Very Relevant). The instrument was designed to measure four latent variables, namely: assessment errors, measurement errors, teacher quality, and madrasah performance.

The assessment error variable comprises 11 items, adopted and adapted from Jeschke et al. (2019). These items are grouped into several aspects: (1) subjectivity of assessors, (2) incon-

sistency in scoring, (3) bias caused by a lack of understanding of the instrument, and (4) inadequate documentation or evidence review during the assessment process. Each aspect reflects an indicator that reveals how accreditation assessors may inadvertently introduce error in evaluating madrasa performance.

The measurement error variable consists of seven items, referring to the theoretical framework developed by Mardapi (2017). The aspects measured include: (1) inaccuracy of instruments in representing constructs, (2) misalignment of indicators with intended competencies, (3) instrument ambiguity or vagueness, and (4) technical or procedural flaws in the assessment process. These indicators were formulated to identify potential weaknesses in the validity and reliability of the assessment tools used during accreditation.

The teacher quality variable is measured using seven items adopted from a questionnaire developed by Mailool et al. (2020). The instrument covers key dimensions of teacher competence: (1) pedagogical competence, (2) professional competence, (3) personal competence, and (4) social competence. Each item was constructed to assess the extent to which teachers in madrasas demonstrate these competencies in their teaching practices and professional behaviour.

The madrasah performance variable is assessed through objective data, namely the 2021 accreditation scores officially obtained by each madrasa included in the study. These scores reflect an aggregate evaluation of various aspects of school quality, including the curriculum, learning process, teacher quality, facilities, graduate outcomes, and school management, as determined by the national accreditation agency.

To ensure the validity of the instruments, a content validation process was carried out through expert judgment. Five experts were involved in this process, consisting of four measurement experts and one educational assessment expert. Each item was rated on a 4-point scale and accompanied by qualitative feedback to improve the clarity and relevance of the items. The revised instruments were then evaluated using Aiken's V coefficient to determine the content validity index. According to established benchmarks, items with a validity coefficient greater than 0.70 ($\alpha > 0.70$) were considered valid and acceptable for inclusion in the instrument (Farida & Setiawati, 2021; Retnawati, 2016).

Furthermore, researchers used exploratory factor analysis (EFA) to see the validity of the instrument empirically. The results of EFA showed a Keyser Mayer Oikin/KMO value of 0.820 (> 0.5), Sig. Barlett's Test of Sphericity was 0.000 (< 0.05), Measures of Sampling Adequacy for each item > 0.5, Eigenvalue> 1.0, and Rotated Component Matrix > 0.5. The EFA results show that the instrument has proven empirically valid (Prihono, Lapele, et al., 2022; Retnawati, 2016). Moreover, a confirmatory factor analysis (CFA) was performed to prove that the instrument has a construct that follows the proposed model referring to the recommended model fit criteria (Hair Jr. et al., 1992; Pedhazur & Kerlinger, 1982; Putro & Lee, 2017).

Estimation of instrument reliability uses Cronbach's Alpha formula and construct reliability (CR). The estimated reliability of Cronbach's Alpha is 0.82 ($\alpha > 0.70$), and CR is 0.79 ($\alpha > 0.70$), indicating that the instrument has good measurement consistency (Retnawati, 2016).

Data Collection Techniques

The data collection technique uses a questionnaire instrument. The data collection process was carried out after the researcher obtained approval from the madrasas involved. Furthermore, the researchers obtained consent from each respondent by guaranteeing their confidentiality and privacy. In the data collection process, the researchers also gave each respondent the freedom to withdraw from the research if they felt the information was not to their liking. This was done to respect the respondent's rights.

Data Analysis Technique

Path analysis was conducted to test the research hypothesis using LISREL 8.80 software. This statistical technique was employed to examine the direct and indirect causal relationships

among the variables. The initial test of the model showed a poor fit based on recommended criteria, including p-value > 0.05, Root Mean Square Error of Approximation (RMSEA) ≤ 0.08, Standardized Root Mean Square Residual (SRMR) ≤ 0.08, Comparative Fit Index (CFI) ≥ 0.90, and Tucker-Lewis Index (TLI) ≥ 0.90 (Hair Jr. et al., 1992; Hair Jr. et al., 2010). Specifically, it evaluated: (1) the direct effects of assessment errors and measurement errors on madrasa performance, and (2) the indirect effects of these errors on madrasa performance mediated by teacher quality. The significance of the hypothesized paths was determined using the t-value criterion: if t > 1.96, the effect was considered statistically significant; if t < 1.96, the effect was not significant (Purwanto et al., 2021). The results of this analysis provide empirical insights into how quality assessment mechanisms may impact madrasa performance, particularly in underdeveloped regions.

FINDINGS AND DISCUSSION

Findings

This study aims to highlight potential disparities in madrasa performance assessment in underdeveloped areas of Maluku Province. These disparities may arise from two primary sources: assessment errors (related to the assessors) and measurement errors (related to the assessment instruments). A better understanding of the direct and indirect effects of these factors on *madrasa* performance is essential for improving the equity of the accreditation system.

Descriptive Statistics of Research Variables

To provide an overview of the distribution and tendencies of the main variables analysed, descriptive statistics, including minimum, maximum, mean, and standard deviation, are presented in Table 2. As shown in Table 2, all variables exhibit relatively high mean scores ranging from 73.36 to 74.98, suggesting that respondents generally perceive these aspects positively. However, the standard deviations ranging from 8.57 to 11.40 indicate considerable variation across respondents, especially in terms of teacher quality. The highest variability in teacher quality responses may reflect differences in teaching competencies, pedagogical skills, or access to professional development across schools in different underdeveloped districts. These descriptive results provide a strong initial context for further inferential analysis using path analysis.

Table 2. Descriptive Statistics of Research Variables

Variables	Minimum	Maximum	Mean	Standard deviation
Assessment error	45.45	95.45	74.51	8.57
Measurement error	46.43	96.43	73.36	9.68
Teacher quality	42.86	100.00	74.98	11.40
Madrasa performance	44.44	97.22	74.81	9.95

Identification of Assessment Inequality

To identify inequality in the madrasa performance assessment in underdeveloped areas of Maluku Province, hypothesis testing was conducted using path analysis. This analysis aimed to examine the direct and indirect effects of assessment error and measurement error on madrasa performance, with teacher quality acting as a mediating variable. The structural model was constructed based on the conceptual framework previously developed in the methods section.

In the initial model test, the results indicated that the model did not meet the recommended goodness-of-fit thresholds. Specifically, fit indices such as RMSEA (0.13), SRMR (0.089), CFI (0.73), and TLI (0.68) fell outside acceptable ranges. These values suggested that the proposed relationships among variables did not adequately reflect the empirical data. To address this issue, model modifications were performed using the modification indices provided by LISREL, along with theoretical justification. Modifications included the removal of non-

significant paths, allowing covariances between related error terms, and excluding indicators with low factor loadings. After these refinements, the revised model showed substantial improvements, with all fit indices falling within the acceptable range. The model fit results for the revised model are presented in Table 3.

Table 3. Model Fit of Inequality Assessment of Underdeveloped Madrasa Performance

Criteria	Recommended Value	Acquisition	Information
Chi-Square with p-value	> 0.05	1.53 ; p = 0.47	fit
RMSEA	< 0.08	0.000	fit
SRMR	< 0.08	0.030	fit
GFI	> 0.90	0.96	fit
NFIs	> 0.90	0.98	fit
CFI	> 0.90	1.00	fit
TLI	> 0.90	0.92	fit

Table 3 demonstrates that the modified model meets the recommended goodness-of-fit criteria, indicating that the revised model effectively captures the relationships between assessment errors, measurement errors, teacher quality, and madrasa performance. Figure 3 shows the standardized solution diagram of the final path model used to test the significance of the relationships among the variables.

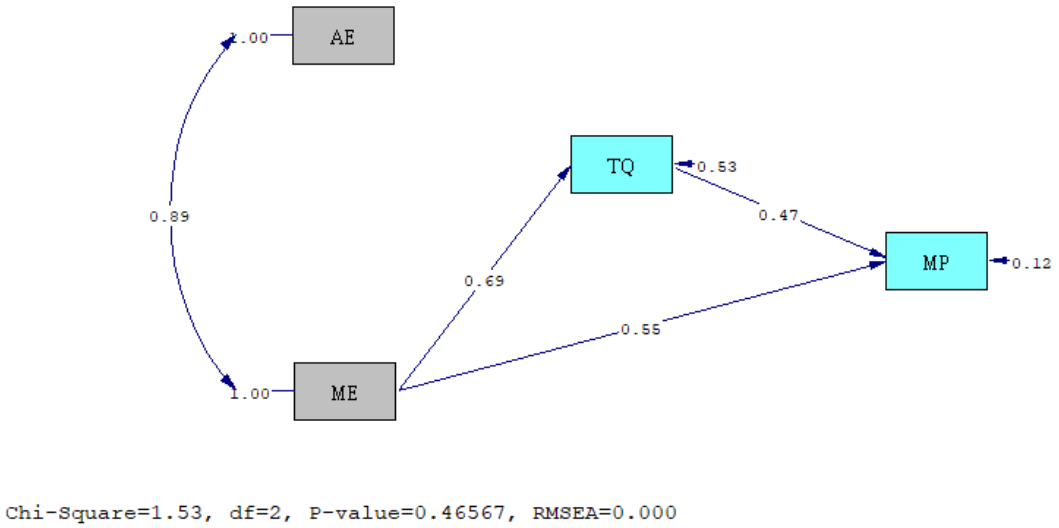

Figure 3. Final Path Model Based on Standardized Solution

Figure 3 presents the final model based on the standardized solution, which reinforces these conclusions by illustrating the structural relationships among variables with standardized path coefficients. In this model, measurement error has a significant influence on teacher quality ($\beta = 0.69$), which in turn positively affects madrasa performance ($\beta = 0.47$). Moreover, teacher quality also has a direct and significant effect on performance, reinforcing its dual role as both a mediator and a direct predictor of performance. Interestingly, a strong covariance ($\beta = 0.89$) was identified between assessment error and measurement error, suggesting that areas prone to subjective assessment biases also tend to face weaknesses in the measurement tools used. Although assessment error does not significantly influence madrasa performance, either directly or indirectly, its correlation with measurement error implies a potential reinforcing mechanism that may exacerbate disparities in assessment outcomes.

The explanatory power of the model is demonstrated through the R-squared values. The R^2 for teacher quality is 0.47, indicating that 47% of its variance is explained by measurement error. For the madrasa performance, the R^2 is 0.53, indicating that more than half of the variation in performance is accounted for by the combined effects of teacher quality and measurement error.

These findings emphasize the central role of valid and reliable measurement instruments in ensuring fair assessments in underdeveloped educational contexts.

Hypothesis testing results further support these conclusions. The direct effect of measurement error on madrasa performance was statistically significant, with a t-value of 4.86 (> 1.96), leading to the acceptance of H2. Similarly, the indirect effect of measurement error on performance through teacher quality was also significant (t-value = 4.16), supporting H4. In contrast, the direct and indirect paths involving assessment error were not statistically significant, resulting in the rejection of H1 and H3.

Table 4. Results of Madrasa Performance Inequality Hypothesis Testing

Effect	Hypothesis	Variable	Correlation Coefficient	Total Effects	Direct Effects	Indirect Effects
Direct	H_2	ME → MP	0.88	0.88	0.88	0
Indirect	H_4	ME → TQ → MP	0.88	0.32	0	0.32

Table 4 shows that the total effect of measurement error on madrasa performance is 0.88 (88%), consisting of a direct effect of 0.55 and an indirect effect of 0.32 mediated by teacher quality. This suggests that measurement error is the most substantial contributor to disparities in performance assessments, while the remaining 12% is explained by other factors not included in the model. These results underscore the importance of improving assessment instruments and strengthening teacher quality as strategic levers to reduce inequality in madrasa performance evaluations, particularly in underdeveloped areas.

Discussion

This study aimed to highlight the potential inequity in the performance assessment of madrasas in underdeveloped areas of Maluku Province. The findings revealed that inequality in performance assessment is primarily influenced by measurement errors related to the instruments used, both directly and indirectly through the mediating role of teacher quality. In contrast, assessment errors related to assessor subjectivity were found to have no significant impact on madrasa performance, either directly or indirectly.

From a descriptive standpoint, the gender distribution of respondents showed a dominance of male teachers (61.1%). This trend aligns with previous studies (Nasir & Mujiati, 2020; Vania et al., 2021), which suggest that geographical isolation, access barriers, and security issues discourage female graduates from teaching in remote areas. This demographic imbalance raises concerns about long-term teacher availability, especially as 26.67% of current teachers are aged between 25–36, potentially leading to a generational gap in the coming decade. Additionally, 12.22% of respondents held only diploma-level qualifications, which is below the national teacher standard requiring at least a bachelor's degree (Koswara & Rasto, 2016) highlighting a persistent challenge in ensuring teacher professionalism in disadvantaged regions.

The central finding of this study is the significant influence of measurement error on performance assessment inequality. According to Mardapi (2017), measurement errors can stem from instruments that are either too difficult or too easy, or from inconsistencies in scoring practices. These errors contribute to systematic bias in assessments, particularly when standardized instruments fail to reflect the socio-educational context of madrasas in underdeveloped regions (Prihono et al., 2023). This supports theoretical assertions by Feldt and Qualls (1998) and Kolen et al. (1992) that educational measurement is inherently indirect and vulnerable to both the physical and psychological conditions of respondents and raters, which can fluctuate over time.

Conversely, assessment errors linked to assessor subjectivity were found to be insignificant. This result may be explained by the professional competence of the assessors. As regulated, assessors are required to adhere to six principles: objectivity, comprehensiveness, fairness, transparency, accountability, and professionalism. These principles help minimise common rater

biases, such as halo effects (Thakkar & Lohiya, 2021), stereotyping (Lucka et al., 2021; Tse & Tung, 2020), linearity, severity, and central tendency errors (Dossanova et al., 2021; Wibowo et al., 2022). The presence of structured tools, rigorous training, and a clear assessment framework appears to reduce variance caused by rater behaviour.

This study fills a research gap identified in the introduction, namely, the limited empirical evidence on how structural measurement factors contribute to assessment inequality in marginalised educational settings. While previous studies have discussed disparities in school infrastructure, access, and teacher distribution, few have systematically investigated how the design and implementation of assessment instruments themselves may contribute to inequality.

Theoretically, this study contributes to the literature by emphasising that instrument validity and contextual appropriateness are as crucial as rater objectivity in achieving fair performance assessments. Practically, the findings underscore the need for context-sensitive instruments that reflect the realities of madrasas in underdeveloped areas. The measurement tools should be developed based on a three-dimensional policy framework: macro (organisational structures and policies), meso (resource behaviour such as leadership and team dynamics), and micro (individual psychological factors like motivation and job satisfaction), as discussed by Robbins and Judge (2009) and recommended in policy development frameworks (Anita & Dwiningrum, 2022; Kawuryan et al., 2021; Wakano et al., 2024).

To minimize measurement errors, instruments must be tailored to reflect local socio-organisational and cultural contexts. For instance, madrasas in underdeveloped areas often differ substantially in governance, teaching capacity, and community expectations compared to urban schools. This calls for an adaptive measurement strategy that includes pilot testing, localised validation, and continuous refinement to ensure both reliability and fairness.

However, this study is not without limitations. First, the sample was limited to 18 madrasas in underdeveloped areas of Maluku Province, which may affect the generalizability of the findings. Second, while structural relationships were established through path analysis, qualitative insights from assessors or policymakers were not included, which may have enriched interpretation. Future research should consider a mixed methods approach to capture both the statistical and experiential dimensions of assessment inequality. Additionally, expanding the model to include other potential mediators, such as leadership effectiveness, school infrastructure, or community involvement, may provide a more comprehensive picture.

This study reveals that inequality in madrasa performance assessments is primarily rooted in instrument-related measurement errors, not assessor subjectivity. These findings advocate for contextual reform in assessment practices, especially in remote and underdeveloped educational regions. Addressing these disparities is not only a technical necessity but also a matter of educational justice.

CONCLUSION

The results of this study confirm that inequality in madrasa performance assessment in underdeveloped areas is significantly influenced by measurement errors related to the use of standardized instruments. These instruments, applied uniformly across developed, developing, and underdeveloped regions, fail to reflect the unique contextual realities faced by madrasas in remote or disadvantaged areas. In contrast, assessment errors stemming from assessor subjectivity did not significantly affect madrasa performance, which may be attributed to the high level of professionalism, objectivity, and training of accreditation teams. Moreover, the study found that teacher quality plays a mediating role in the relationship between measurement error and performance, highlighting the importance of human resource factors in educational evaluation. However, this finding also reveals a deeper structural issue: that uniform tools in diverse educational landscapes may inadvertently disadvantage institutions in less developed regions. This highlights the need for context-sensitive assessment frameworks that acknowledge local conditions and constraints.

Despite these important findings, this study has certain limitations. The research was conducted in a specific geographic and administrative context, underdeveloped areas of Maluku Province, which may limit the generalizability of the results to other regions. Additionally, the study primarily focused on teacher quality as a mediating variable, without incorporating other potentially influential factors, such as infrastructure, curriculum implementation, leadership, or socioeconomic environment. Future studies should therefore adopt a broader scope by integrating both physical and psychological dimensions of educational quality, such as management systems, facilities, student learning outcomes, and school leadership. To promote fair and valid assessment practices, policymakers and practitioners should consider developing performance measurement instruments that are grounded in macro, meso, and micro perspectives. These tools should account for organizational dynamics, teacher engagement, and individual psychological conditions, ensuring that accreditation processes contribute to equitable and just educational development in all regions.

ACKNOWLEDGMENT

The authors express their gratitude to the respondents from the 18 madrasas across Maluku Province, whose participation was instrumental to the successful conduct of this study. The authors also formally acknowledge the Ministry of Religious Affairs for its financial support, provided through the Institut Agama Islam Negeri Ambon (now Universitas Islam Negeri Abdul Muthalib Sangadji Ambon), which contributed significantly to the completion of this research.

DISCLOSURE STATEMENT

The authors declare that they have no conflicts of interest to disclose.

FUNDING STATEMENT

This work was supported by the Ministry of Religious Affairs through its program for enhancing research quality at the State Islamic Institute of Ambon (Universitas Islam Negeri Abdul Muthalib Sangadji Ambon), under Grant Number 28 of 2023.

ETHICS APPROVAL

This study complied with ethical standards and data privacy regulations. All data were used solely for research, kept confidential, and handled securely.

REFERENCES

Anita, A., & Dwiningrum, S. I. A. (2022). Digitalisasi dan ketimpangan pendidikan: Studi kasus terhadap guru sekolah dasar di Kecamatan Baraka. *Jurnal Pendidikan dan Kebudayaan*, *7*(1), 1–12. https://doi.org/10.24832/jpnk.v7i1.2509

Bakhri, A. (2015). Tantangan pendidikan agama Islam di madrasah pada era globalisasi. *Madaniyah*, *5*(1), 63–86. https://journal.stitpemalang.ac.id/index.php/madaniyah/article/download/66/48

Bida, O. (2021). Kebijakan desentralisasi dan disparitas pendidikan lokal di Provinsi Sulawesi Tenggara. *Jurnal Ilmu Administrasi Negara ASIAN (Asosiasi Ilmuwan Administrasi Negara)*, *9*(1), 228–248. https://doi.org/10.47828/jianaasian.v9i1.59

Dossanova, N., Abdimanuly, O., Maulenov, A., Karbozov, Y., & Matbek, N. (2021). Cultural interaction in the works of M. Zhumabayev as the basis for the development of a national literary tradition. *International Journal of Society, Culture & Language*, *9*(2, Themed issue on Modern Realities of National Languages of CIS Countries), 41–53. https://www.ijscl.com/article_245175.html

Dwiningrum, S. I. A. (2007). Desentralilsasi pendidikan dan ketimpangan mutu pada tingkat satuan pendidikan. In *Makalah Simposium Nasional 2007, Jakarta 26-27 Juli 2007*. Pusat Penelitian Kebijakan dan Inovasi Pendidikan, Balitbang, Depdiknas RI.

Farida, S., & Setiawati, F. A. (2021). Developing assessment instruments of debate practice in Indonesian language learning. *REID (Research and Evaluation in Education)*, *7*(2), 145–155. https://doi.org/10.21831/reid.v7i2.43338

Feldt, L. S., & Qualls, A. L. (1998). Approximating scale score standard error of measurement from the raw score standard error. *Applied Measurement in Education*, *11*(2), 159–177. https://doi.org/10.1207/s15324818ame1102_3

Hair Jr., J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1992). *Multivariate analysis with readings* (3rd ed.). McMillan Publishing Company.

Hair Jr., J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis. In *Multivariate data analysis* (7th ed.) (p. 785). Prentice Hall.

Hasanah, E. (2021). Best practice penjaminan mutu lulusan berbasis IASP 2020 di sekolah menengah kejuruan. *JAMP: Jurnal Administrasi dan Manajemen Pendidikan*, *4*(2), 178–186. https://journal-fip.um.ac.id/index.php/jamp/article/view/1683

Hasanah, N., Latuapo, R., Prihono, E. W., & Ndayizeye, O. (2025). An analysis of determinant factors in the management of madrasah ibtidaiyah in the archipelago region. *Nidhomul Haq: Jurnal Manajemen Pendidikan Islam*, *10*(1), 279–292. https://doi.org/10.31538/ndhq.v10i1.36

Hermawansyah, H. (2021). Manajemen lembaga pendidikan sekolah berbasis digitalisasi di era Covid-19. *Fitrah: Jurnal Studi Pendidikan*, *12*(1), 27–46. https://ejournal.stitbima.ac.id/index.php/fitrah/article/view/320

Huda, K. (2016). Problematika madrasah dalam meningkatkan mutu pendidikan islam. *Dinamika Penelitian: Media Komunikasi Penelitian Sosial Keagamaan*, *16*(2), 309–336. https://doi.org/10.21274/dinamika.2016.16.2.309-336

Jeschke, C., Kuhn, C., Lindmeier, A., Zlatkin-Troitschanskaia, O., Saas, H., & Heinze, A. (2019). Performance assessment to investigate the domain specificity of instructional skills among pre-service and in-service teachers of mathematics and economics. *British Journal of Educational Psychology*, *89*(3), 538–550. https://doi.org/10.1111/bjep.12277

Jumono, S., Baskara, I., Abdurahman, A., & Mala, C. M. F. (2021). The dynamics of economic growth in underdeveloped regions: A case study in Indonesia. *The Journal of Asian Finance, Economics and Business*, *8*(4), 643–651. https://doi.org/10.13106/jafeb.2021.vol8.no4.0643

Kawuryan, S. P., Sayuti, S. A., & Dwiningrum, S. I. A. (2021). Teachers quality and educational equality achievements in Indonesia. *International Journal of Instruction*, *14*(2), 811–830. https://doi.org/10.29333/iji.2021.14245a

Khilji, G., Jogezai, N. A., Bibi, N., & Baloch, F. A. (2022). Understanding the effects of educational decentralization through the perspectives of education managers and teachers. *International Journal of Educational Management*, *36*(7), 1206–1220. https://doi.org/10.1108/IJEM-05-2021-0181

Kolen, M. J., Hanson, B. A., & Brennan, R. L. (1992). Conditional standard errors of measurement for scale scores. *Journal of Educational Measurement*, *29*(4), 285–307. https://doi.org/10.1111/j.1745-3984.1992.tb00378.x

Koswara, K., & Rasto, R. (2016). Kompetensi dan kinerja guru berdasarkan sertifikasi profesi. *Jurnal Pendidikan Manajemen Perkantoran (JPManper)*, *1*(1), 61–71. https://ejournal.upi.edu/index.php/jpmanper/article/view/3269

Lucka, N. S., Caldieraro, F., & Zanini, M. T. (2021). The influence of gender stereotyping and issue advocacy on consumer sentiment. *Marketing Intelligence & Planning, 39*(6), 777-791. https://doi.org/10.1108/MIP-11-2020-0469

Mailool, J., Kartowagiran, B., Retnowati, T. H., Wening, S., & Putranta, H. (2020). The effects of principal's decision-making, organizational commitment and school climate on teacher performance in vocational high school based on teacher perceptions. *European Journal of Educational Research*, *9*(4), 1675–1687. https://doi.org/10.12973/eu-jer.9.4.1675

Mardapi, D. (2017). *Pengukuran penilaian dan evaluasi pendidikan*. Parama Publishing.

Nasir, N., & Mujiati, M. (2020). Pengelolaan tenaga pendidik (guru) di rural area. *Didaktis: Jurnal Pendidikan dan Ilmu Pengetahuan*, *20*(1), 37-53. https://doi.org/10.30651/didaktis.v20i1.4383

Pedhazur, E. J., & Kerlinger, F. N. (1982). *Multiple regression in behavioral research*. Holt, Rinehart, and Winston.

Presidential Regulation No. 63 of 2020 concerning the Designation of Underdeveloped Regions for the 2020–2024 Period. (2020). Republic of Indonesia.

Prihono, E. W., Haryanto, H., & Munadi, S. (2025). Accuracy of fuzzy clustering based on school accreditation data: Reflection and evaluation for improving education quality. *TEM Journal*, *14*(2), 1715–1724. https://doi.org/10.18421/TEM142-69

Prihono, E. W., Lapele, F., Jumaeda, S., Sukadari, S., & Nurjanah, S. (2022). EFA of pedagogic competence instrument to measure teacher performance. In *5th International Conference on Current Issues in Education (ICCIE 2021)*, 323–328. https://doi.org/10.2991/assehr.k.220129.059

Prihono, E. W., Retnawati, H., Lapele, F., & Waluyo, W. B. (2022). The quality of Indonesian teachers in the digital era: A meta-analysis. *Jurnal Penelitian dan Evaluasi Pendidikan*, *26*(2), 186–200. https://doi.org/10.21831/pep.v26i2.52318

Prihono, E. W., Waluyo, W. B., Lapele, F., Retnawati, H., & Setiawan, C. (2023). Analyzing madrasah teacher performance assessment in underdeveloped areas: A phenomenological study in Maluku Province. *Indonesian Journal of Islamic Education Studies (IJIES)*, *6*(1), 1–16. https://doi.org/10.33367/ijies.v6i1.3494

Purwanto, A., Asbari, M., & Santoso, T. I. (2021). Analisis data penelitian sosial dan manajemen: Perbandingan hasil antara Amos, SmartPLS, WarpPLS, dan SPSS untuk jumlah sampel medium. *International Journal of Social and Management Studies*, *2*(4), 43–53. https://doi.org/10.5555/ijosmas.v2i4.50

Putro, N. H. P. S., & Lee, J. (2017). The structural equation modeling of reading interest psycho-behavioural constructs: How are they related across different modes of reading? *REID (Research and Evaluation in Education)*, *3*(1), 50–63. https://doi.org/10.21831/reid.v3i1.13530

Retnawati, H. (2016). *Validitas reliabilitas dan karakteristik butir*. Parama Publishing.

Robbins, S. P., & Judge, T. (2009). *Organizational behavior*. Pearson South Africa.

Sari, E. A., Saragih, M. T. B., Shariati, I. A., Sofyan, S., Al Baihaqi, R., & Nooraeni, R. (2020). Klasifikasi kabupaten tertinggal di kawasan timur Indonesia dengan support Vector Machine. *JIKO (Jurnal Informatika dan Komputer)*, *3*(3), 188–195. https://doi.org/10.33387/jiko.v3i3.2364

Thakkar, A., & Lohiya, R. (2021). A survey on intrusion detection system: Feature selection, model, performance measures, application perspective, challenges, and future research directions. *Artificial Intelligence Review*, 1–111. https://doi.org/10.1007/s10462-021-10037-9

Tse, W. T. S., & Tung, V. W. S. (2020). Assessing explicit and implicit stereotypes in tourism: Self-reports and implicit association test. *Journal of Sustainable Tourism*, 31(2), 460-482. https://doi.org/10.1080/09669582.2020.1860995

Vania, A. S., Septianingrum, A. D., Suhandi, A. M., & Prihantini, P. (2021). Revitalisasi peran guru dalam pengelolaan kelas di daerah terdepan, terluar, dan tertinggal (3T) pada era Revolusi Industri 4.0. *Jurnal Basicedu*, *5*(6), 5142–5150. https://doi.org/10.31004/basicedu.v5i6.1587

Wakano, A., Latuapo, R., Lapele, F., & Prihono, E. W. (2024). Measuring teacher personality competence: Era 5.0 challenges for teachers in Indonesian archipelago. *AL-ISHLAH: Jurnal Pendidikan*, *16*(3), 4047–4057. https://doi.org/10.35445/alishlah.v16i3.4713

Wardi, M., Ismail, I., & Makki, A. (2019). Perbandingan pendidikan; pemahaman simbolis dan substantif PAI di madrasah dan PAI di sekolah umum. *Nidhomul Haq: Jurnal Manajemen Pendidikan Islam*, *4*(1), 23–33. https://doi.org/10.31538/ndh.v4i1.104

Wibowo, A. A., Siregar, M., & Mulyo, M. T. (2022). The pattern of internalization of aswaja an-nahdliyah character values (Analysis study of madrasah aliyah with Islamic education background). *Journal Research of Social, Science, Economics, and Management*, *1*(9), 1440–1452. https://doi.org/10.36418/jrssem.v1i9.154