

Rethinking Diagnostic Assessment Through Culturally Embedded AI for Meaningful Learning

Benny Anggara^{1,2}, Sugiman¹, Elly Arliani¹

¹Department of Mathematics Education, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

²Department of Mathematics Education, Universitas Sindang Kasih Majalengka, West Java, Indonesia

Correspondence: Sugiman, Department of Mathematics Education, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia.

Received: August 6, 2026

Accepted: September 28, 2026

Online Published: October 8, 2026

doi:10.11114/jets.v15i1.9159

URL: <https://doi.org/10.11114/jets.v15i1.9159>

Abstract

Advances in artificial intelligence have accelerated the transformation of educational assessment, but most innovations remain focused on automating scoring rather than generating meaningful evidence of learning. This study aims to develop and validate the Meaningful Diagnostic Assessment Framework (MDAF), which integrates culturally contextualised diagnostic tasks, AI-based analysis, and pedagogical interpretation to generate meaningful evidence of learning in mathematics education. The study employed a controlled quasi-experimental design involving 178 junior high school students in Indonesia. Data were analysed using Structural Equation Modelling (SEM), mediation analysis, and tests of agreement between AI-generated diagnoses and expert assessments. The results indicate positive associations between the MDAF and symbolic mathematical performance and student engagement. Cultural Embeddedness Perception partially mediated these relationships, while AI-generated diagnoses showed high agreement with expert assessments. These findings suggest that AI's primary value in assessment lies not in automatically generating scores, but in transforming student responses into meaningful evidence of learning that supports pedagogical decision-making. This study offers a new paradigm for diagnostic assessment by shifting the focus from “assessment as measurement” to “assessment as the generation of meaningful learning evidence,” thereby providing a conceptual foundation for developing digital assessments that are more adaptive, transparent, and oriented toward the learning process.

Keywords: Meaningful Diagnostic Assessment Framework (MDAF), meaningful learning evidence, artificial intelligence, cultural embeddedness, symbolic mathematical reasoning, adaptive assessment

1. Introduction

Advances in digital technology have given education systems an unprecedented ability to record, analyse, and utilise learning data at scale. However, this capability has not been accompanied by commensurate changes in how schools document students' learning processes (Baker et al., 2021; Kurniawan et al., 2026; Mabalane Jr., 2025). Various educational agendas, including Sustainable Development Goal (SDG) 4 and the OECD framework through the Programme for International Student Assessment (PISA), prioritize the quality of learning, which is no longer measured solely by academic achievement but by the education system's ability to provide information that explains how students construct knowledge, develop reasoning, and overcome learning difficulties (Liu et al., 2021; Meyer et al., 2025). This paradox highlights that the primary challenge in education today is no longer merely introducing technology into the classroom, but transforming assessment into a source of learning evidence that supports more meaningful and adaptive learning.

Developments in assessment theory indicate a shift from the “assessment of learning” paradigm toward “assessment for learning,” which uses assessment results to inform instructional decisions. This shift has spurred the development of diagnostic assessments designed to reveal thought processes, identify misconceptions, and map students' learning characteristics more comprehensively (Graven & Venkat, 2021; Zhang, Zhao, et al., 2025). Open-ended assessment is increasingly being adopted because it allows students to demonstrate strategies, representations, arguments, and reasoning that closed-ended questions cannot reveal. This information yields richer evidence for learning, enabling teachers to design more targeted interventions (Chen et al., 2026). Most diagnostic assessments still rely on manual interpretation, which is time-consuming, depends on rater consistency, and cannot efficiently or systematically generate student reasoning profiles at scale.

Advances in artificial intelligence have opened new opportunities in educational assessment by automating scoring, answer classification, learning performance prediction, and rapid analysis of learning data. Various studies indicate that this technology can improve the efficiency and consistency of the evaluation process (Mahamad et al., 2025; Sacristán, 2024; Zhai et al., 2021). These advancements have not yet fully transformed the assessment paradigm because most AI implementations still prioritise scoring accuracy over understanding the reasoning processes underlying students' responses (Aldaleel, 2025). The diagnostic information generated often remains limited to whether an answer is correct or incorrect and thus fails to illustrate how knowledge is constructed (McNamara et al., 2023). Cultural context offers a different perspective because situations closely aligned with students' experiences can elicit more authentic thinking strategies, reveal how they construct meaning around concepts, and yield responses that more accurately reflect their actual abilities (Chang & Viesca, 2022; Prahmana, 2022). Integrating AI-based analysis with cultural context can produce assessments that are not only efficient but also more meaningful in revealing mathematical thinking processes.

This study is based on the view that the main challenge of diagnostic assessment lies not only in the speed or accuracy of scoring, but also in generating information that truly represents students' learning processes. Based on this perspective, this study proposes a framework for Culturally Embedded AI-Based Diagnostic Assessment that integrates culturally contextual, open-ended tasks with AI-based response analysis to generate a more comprehensive profile of mathematical reasoning. This approach offers a conceptual contribution by redefining the function of diagnostic assessment as a provider of meaningful learning evidence; a methodological contribution through the development of an assessment framework that combines cultural context, AI analysis, and reasoning modelling; and a practical contribution in the form of diagnostic information that teachers can utilise to design more adaptive instruction. Based on this framework, this study aims to answer two main questions: (1) How can a culturally embedded AI-based diagnostic assessment be developed to support meaningful learning? Moreover, (2) To what extent does the proposed assessment provide valid, accurate, and educationally meaningful diagnostic evidence for supporting mathematics learning?

1.1 From Measuring Learning to Generating Meaningful Learning Evidence

Developments in assessment theory indicate that assessment's function has fundamentally transformed as the educational field's perspective on the learning process has shifted. In the early paradigm, assessment was framed as "assessment of learning," focused on measuring outcomes through scoring, classification, and reporting (Confrey & Shah, 2021; X. Wu et al., 2022). This approach provided information on competency mastery but offered little insight into how students construct knowledge or why they experience learning difficulties. A shift in perspective on the nature of learning subsequently gave rise to the "assessment for learning" paradigm, which uses assessment results to provide feedback and continuously improve the learning process. Further developments led to "assessment as learning," which positions students as active agents who reflect on, monitor, and regulate their own learning processes (Alfageh et al., 2024). This evolution demonstrates that assessment is no longer understood as a final activity to evaluate learning outcomes but rather as a process of generating evidence of learning that explains students' progress in understanding, thinking strategies, and learning needs (Bento Miguens et al., 2024; Shishigu et al., 2026). This perspective reinforces the development of diagnostic assessment focused on identifying cognitive processes, patterns of reasoning, and misconceptions as the foundation for designing more adaptive and evidence-based instruction. The primary value of modern assessment no longer lies in producing accurate scores but in providing meaningful information to support pedagogical decision-making and improve learning quality.

1.2 Diagnostic Assessment for Revealing Mathematical Reasoning

Modern mathematics education no longer views learning success as the ability to arrive at the correct answer, but rather as the ability to construct the reasoning that underlies the formation of mathematical concepts. This perspective positions mathematical reasoning as the core of mathematical activity because, through this process, students recognise relationships, construct logical arguments, develop generalisations, and connect various representations to generate mathematical meaning (Kong & Wang, 2024; Setiyani et al., 2022). This reasoning becomes particularly important during the early algebraic reasoning phase, when students begin to transition from arithmetic procedures to using symbols to represent relationships between quantities (Ünal et al., 2023). Conventional assessments that focus solely on the accuracy of the final answer cannot adequately capture the complexity of this process. The open-ended assessment approach provides greater opportunities to uncover the strategies, representations, arguments, and variations in thinking that students use to solve a problem, while construct-based assessment allows each response to be interpreted based on theoretical constructs that represent the dimensions of mathematical reasoning (Ardianto et al., 2025; Kohen & Nitzan-Tamar, 2022; Zhan et al., 2024). This type of information provides a much richer profile of reasoning than achievement scores alone, as it reveals each student's conceptual strengths, misconceptions, and learning needs. The main challenge remains interpreting open-ended responses, which is generally done manually; consequently, it is time-consuming, prone to inter-rater variability in scoring, and difficult to implement consistently at scale (Engelhard & Wang, 2024). This suggests that the current challenge in mathematics assessment is no longer generating more measurements but producing more accurate, consistent, and meaningful reasoning profiles to support instructional decision-making.

1.3 Rethinking Diagnostic Assessment Through Cultural Context and Artificial Intelligence

Developments in diagnostic assessment indicate that the quality of the information generated depends not only on the accuracy of the analysis but also on the assessment tasks' ability to elicit authentic reasoning. The culturally responsive assessment perspective views cultural experiences as an integral part of the knowledge-construction process because contexts closely tied to students' lives enable them to construct meaning, connect experiences with mathematical concepts, and reveal thinking strategies more naturally (Alam & Mohanty, 2023; Andrews et al., 2023; Löfström et al., 2026; Tanase & Kayaalp, 2024). Ethnomathematics research reinforces this view by demonstrating that cultural activities, local wisdom, and everyday practices contain mathematical structures that can serve as both learning contexts and assessment contexts (Eglash, 2023; Sunzuma & Umbara, 2025). When diagnostic tasks are developed based on students' familiar experiences, the resulting responses not only demonstrate the ability to arrive at an answer but also reveal how mathematical relationships are constructed, how representations are chosen, and how reasoning develops during the problem-solving process (Huang et al., 2024; Zhang, Zhang, et al., 2025). Assessments rooted in cultural contexts thus yield more authentic evidence of learning because they capture thought to process in situations meaningful to students.

Assessing authentic reasoning requires an analytical system that can consistently interpret open-ended responses in a construct-based manner. Advances in artificial intelligence offer such opportunities through the use of Natural Language Processing to extract semantic meaning from students' answers, construct-aligned scoring to link responses to the measured reasoning constructs, and automated reasoning analysis and learning analytics to systematically identify thinking patterns, problem-solving strategies, and tendencies toward misconceptions (Bayaga, 2025; Richardo et al., 2023). Advances in explainable artificial intelligence further improve transparency by providing a traceable, accountable basis for interpretation (Han et al., 2025; J. Wu et al., 2025). Integrating cultural context with AI-based analysis introduces a new paradigm for diagnostic assessment that is no longer score-oriented but instead focuses on building meaningful evidence of learning that reflects how students construct knowledge (Luo, 2023; Putra et al., 2026). From this perspective, the study proposes the Meaningful Diagnostic Assessment Framework (MDAF). This conceptual framework integrates culturally embedded assessment and AI-supported reasoning analysis to generate a comprehensive reasoning profile that supports adaptive and more meaningful learning.

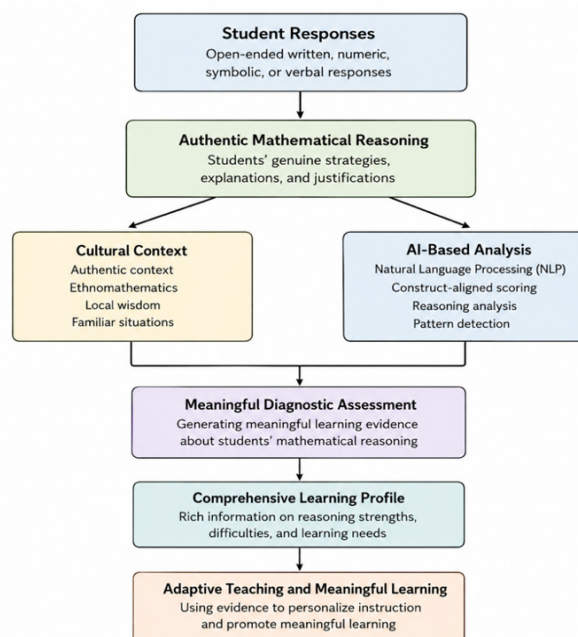


Figure 1. Meaningful Diagnostic Assessment Framework

2. Method

2.1 Study Design

This study employed a quantitative controlled quasi-experimental design to examine the implementation of the Meaningful Diagnostic Assessment Framework (MDAF) in mathematics learning. The framework was developed on the assumption that diagnostic assessments need to generate meaningful evidence of learning by integrating culturally contextual, open-ended tasks with artificial intelligence-assisted response analysis. The study compared two conditions: an experimental condition using the MDAF-based diagnostic assessment and a comparison condition using a

conventional digital diagnostic assessment. Participants were drawn from existing classes rather than individually randomized, allowing the study to be conducted within the regular school setting. Students first completed a baseline assessment to establish their initial levels of symbolic mathematical performance and student engagement before the respective assessment conditions were implemented. The research process followed four stages: designing culturally based diagnostic tasks, developing an AI-based response analysis mechanism, implementing the two assessment conditions, and analysing relationships among constructs using Structural Equation Modelling (SEM) (Feng & Hancock, 2024; J. Hair & Alamer, 2022). This series of steps was designed to provide empirical evidence on the effectiveness and pedagogical validity of the proposed assessment framework. The implementation took place over five sessions, each lasting approximately one hour.

2.2 Participants and Research Setting

This study involved 178 junior high school students from several schools in West Java Province, Indonesia, that have integrated digital technology into mathematics instruction. Participants were distributed across two conditions: 90 students (50.6%) in the experimental group and 88 students (49.4%) in the comparison group. We selected junior high school students because this level represents a critical transitional phase toward more formal symbolic representations, making it appropriate for evaluating symbolic mathematical performance and student engagement in a technology-based learning environment.

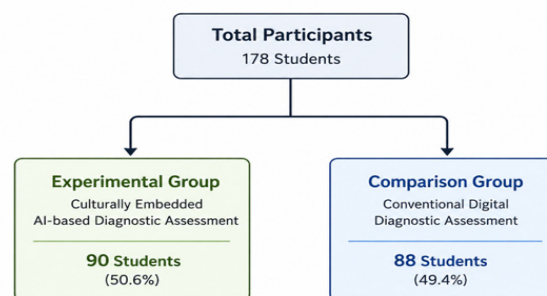


Figure 2. Distribution of Participants Across Experimental Conditions

Participants were selected using purposive sampling at the class level, considering students' initial mathematics ability. We purposively selected existing classes with relatively comparable initial mathematics ability for the experimental and comparison conditions. The selected experimental class had an average initial mathematics ability score of $M = 78$ ($SD = 9.5$), while the comparison class had $M = 79$ ($SD = 9.6$). All research activities were conducted during regular class hours without altering the class structure or the learning process established by the school. Student participation was voluntary, with consent from the school and parents or guardians, and all data were analysed anonymously and used solely for research purposes in accordance with educational research ethics.

2.3 Operationalisation of the Meaningful Diagnostic Assessment Framework

In this study, the MDAF was operationalised as a diagnostic process that transforms students' open-ended responses into interpretable evidence of mathematical learning. The process consisted of five interconnected stages: culturally embedded task input, student response input, AI-supported diagnostic analysis, generation of meaningful learning evidence, and pedagogical decision-making. The operationalisation was designed to ensure that the assessment process captured not only whether students obtained correct answers but also how they represented, reasoned about, and interpreted the mathematical relationships embedded in the tasks.

Figure 3 shows how MDAF transforms students' responses to culturally embedded open-ended tasks into diagnostic and pedagogical evidence. The process begins with the culturally embedded task, followed by collecting students' open-ended responses. GradeLab then analyses these responses through response interpretation, construct-aligned analysis, and reasoning analysis. The resulting diagnostic profiles provide evidence of students' mathematical reasoning, symbolic performance, misconceptions, and engagement, which can subsequently inform targeted feedback, differentiated instruction, and further instructional decision-making. The feedback loop shows that diagnostic evidence can inform subsequent instructional responses and the design of further assessment activities.

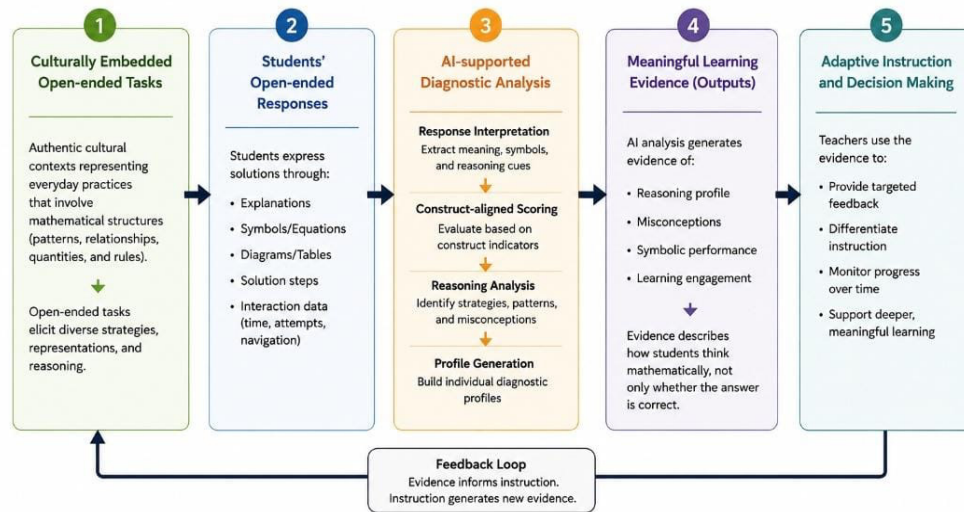


Figure 3. Operational mechanism of the MDAF

The process began with administering culturally embedded open-ended diagnostic tasks. Students' responses constituted the primary input for GradeLab and included their final answers, mathematical expressions or symbols, and written explanations of their reasoning. The tasks were designed to connect mathematical relationships with recognisable cultural artefacts or practices while maintaining a genuine mathematical problem structure. Figure 4 presents a representative task using *Pedati Gede*, a traditional cultural artefact from Cirebon. The task asks students to represent relationships among distances between the wheel centres using symbolic variables and to explain the relationships they used. The figure also illustrates the cultural context, the source of the mathematical quantities, the task's diagnostic focus, and examples of anticipated response patterns.

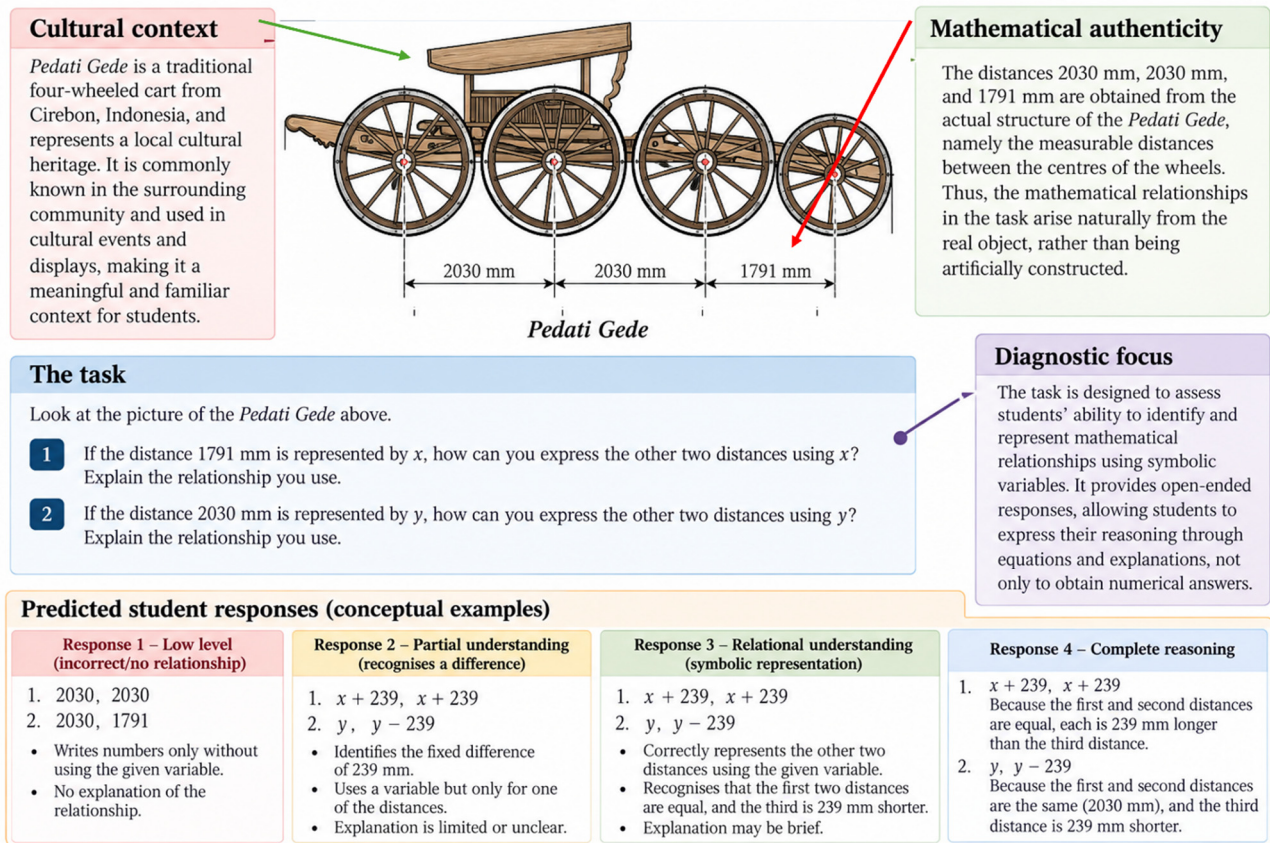


Figure 4. Representative culturally embedded diagnostic task

The task's cultural relevance was established through Pedati Gede as a local cultural artefact, while its mathematical authenticity was grounded in measurable distances derived from the artefact's structure. Thus, the mathematical relationships were embedded in the cultural object rather than artificially added to the task. The open-ended format also enabled students to express different forms of mathematical reasoning and symbolic representation. GradeLab then processed the responses through response interpretation, construct mapping, and reasoning analysis. The system identified relevant mathematical expressions, symbols, explanations, and other semantic information and mapped them onto predefined indicators, particularly mathematical relationships, patterns, and generalisation. The analysis also examined students' strategies and possible misconceptions. These results were organised into an individual diagnostic profile describing students' symbolic representations, reasoning strategies, possible misconceptions, and performance on the targeted constructs.

The diagnostic profiles were then validated against expert evaluations involving five experts. Using the same construct indicators and diagnostic criteria, the experts examined students' responses and the corresponding diagnostic interpretations. We assessed agreement between GradeLab-generated diagnoses and expert evaluations using the Intraclass Correlation Coefficient (ICC) and Cohen's Kappa. The validated diagnostic profiles were subsequently interpreted as meaningful learning evidence. This evidence provided information about students' mathematical representations, reasoning strategies, possible misconceptions, and performance beyond a single correct-or-incorrect score. The resulting evidence was intended to support targeted feedback, differentiated instruction, and subsequent instructional decision-making. Thus, MDAF connects students' responses to AI-supported analysis, validated diagnostic interpretation, and pedagogically meaningful evidence of learning.

2.4 Research Instruments

This study uses three main instruments to evaluate the MDAF's effectiveness: the Meaningful Diagnostic Assessment (MDA), the Student Engagement Questionnaire, and the Cultural Embeddedness Perception Scale. The MDA was developed as a context-based, open-ended diagnostic task to assess symbolic mathematical performance across the domains of relationships, patterns, and generalisation. All responses were analysed using GradeLab through a NLP approach, construct-aligned scoring, and reasoning analysis. The Student Engagement Questionnaire measured students' behavioural, emotional, and cognitive engagement during technology-based learning. The Cultural Embeddedness Perception Scale measured students' perceptions of the meaningfulness, authenticity, and relevance of the cultural context used in the diagnostic tasks. A summary of the characteristics of each instrument is presented in Table 1.

Table 1. Research Constructs and Measurement Specifications

Instrument	Purpose	Construct / Indicators	Format
Meaningful Diagnostic Assessment (MDA)	Assess symbolic mathematical performance through culturally embedded diagnostic tasks	Relationships, Generalization	Patterns, Open-ended tasks
Student Engagement Questionnaire	Measure students' engagement during technology-enhanced mathematics learning	Behavioral, Emotional, Cognitive engagement	Five-point Likert scale
Cultural Embeddedness Perception Scale	Measure students' perceptions of cultural relevance	Familiarity, Meaningfulness, relevance	Authenticity, Contextual, Five-point Likert scale

We ensured the suitability of all instruments through validity and reliability testing before structural analysis. Content validity was confirmed through an expert review by three experts in mathematics education, educational technology, and cultural studies, with Content Validity Index (CVI) values ranging from 0.86 to 0.94, indicating how well the indicators align with the constructs being measured. The questionnaire showed good reliability, with Cronbach's alpha coefficients of 0.87 for Student Engagement and 0.89 for Cultural Embeddedness Perception. The suitability of the measurement model was evaluated using Confirmatory Factor Analysis (CFA), with all factor loadings exceeding 0.60, a Composite Reliability (CR) exceeding 0.80, and an Average Variance Extracted (AVE) exceeding 0.50. These results indicate that all instruments meet the validity and reliability requirements and are therefore suitable for use in SEM analysis.

2.5 Data Analysis

Data analysis was conducted using SEM with Maximum Likelihood Estimation (MLE) in IBM SPSS AMOS 26 to test the conceptual model proposed in this study. The analysis focused on testing the effect of MDA on symbolic mathematical performance and student engagement, as well as the mediating role of Cultural Embeddedness Perception in these relationships. Before conducting the structural analysis, the data were examined in IBM SPSS Statistics 26 to ensure completeness, multivariate normality, and the absence of extreme outliers.

We tested mediation effects using a bootstrapping procedure with 5,000 resamples to estimate indirect effects and their 95% confidence intervals. We considered a mediation effect significant when the bootstrapped confidence interval did

not include zero. We evaluated the accuracy of the AI-based diagnostic system by comparing the diagnostic profiles generated by GradeLab with assessments from mathematics education experts, which served as an external reference. The degree of agreement between the two methods was assessed using the Intraclass Correlation Coefficient (ICC) and Cohen's Kappa (Hair et al., 2021).

3. Results

This section presents empirical findings to evaluate the effectiveness of MDAF in supporting meaningful mathematics learning. The research results are presented in stages, beginning with an evaluation of the structural model's feasibility, testing the framework's impact on symbolic mathematical performance and student engagement, analysing the role of Cultural Embeddedness Perception as a mechanism explaining the framework's effectiveness, and evaluating the alignment between diagnoses generated by the AI system and expert assessments. These results provide empirical evidence of the conceptual validity, pedagogical effectiveness, and measurement accuracy of the proposed framework.

3.1 Empirical Support for the Meaningful Diagnostic Assessment Framework

The MDAF was evaluated to ensure that the proposed conceptual structure adequately represented the empirical data before further analysing the relationships among variables. The evaluation used several goodness-of-fit indices commonly used in SEM to assess the fit between the theoretical model and the research data. Table 2 presents the model evaluation results.

Table 2. Empirical Fit of the MDAF

Fit Index	Recommended Value	Obtained Value	Evaluation
χ^2/df	≤ 3.00	2.14	Good fit
Comparative Fit Index (CFI)	≥ 0.90	0.94	Good fit
Tucker–Lewis Index (TLI)	≥ 0.90	0.93	Good fit
Root Mean Square Error of Approximation (RMSEA)	≤ 0.08	0.061	Good fit
Standardized Root Mean Square Residual (SRMR)	≤ 0.08	0.048	Good fit

The empirical fit indicates that the MDAF provides an adequate fit to the empirical data. The relationship structure formulated within the framework consistently represents the interrelationships among constructs, thereby providing empirical support for the proposed conceptual model. These results indicate that integrating AI-based diagnostic assessment, cultural context, symbolic mathematical performance, and student engagement forms a coherent framework, rather than merely a collection of statistically linked variables.

Empirical support for the structural model provides a strong foundation for continuing to test causal relationships in the next phase. These conditions indicate that direct effects, mediation mechanisms, and the evaluation of AI diagnostic accuracy can be interpreted as part of a single, cohesive conceptual framework. These findings provide empirical support for the proposed MDAF as a framework linking cultural context, AI-supported diagnostic assessment, and meaningful learning evidence.

3.2 Associations Between the MDAF and Learning Outcomes

Direct effect tests examined the relationships between the MDAF and the two main learning-related outcomes: symbolic mathematical performance and student engagement. The analysis focused on whether participation in the MDAF condition was positively associated with these outcomes within the structural model. A summary of the direct effect test results is presented in Table 3.

Table 3. Direct Effects of the MDAF on Learning Outcomes

Outcome Variable	Standardized Effect (β)	p-value	Interpretation
Symbolic Mathematical Performance	0.42	< .001	Significant positive effect
Student Engagement	0.37	< .001	Significant positive effect

The results show significant positive associations between the MDAF and both learning outcomes in the structural model. The standardized coefficients indicate that the MDAF was positively associated with symbolic mathematical performance ($\beta = 0.42$, $p < .001$) and student engagement ($\beta = 0.37$, $p < .001$). These findings provide empirical support for the proposed relationships between the MDAF and the learning-related outcomes, although they should not be interpreted as evidence of causal improvement because the present analysis does not report between-group changes from baseline to post-intervention outcomes. These findings provide preliminary evidence that the MDAF is associated with both mathematical performance and aspects of the learning process relevant to meaningful learning. The positive relationship between symbolic mathematical performance and student engagement further suggests that the framework can generate learning evidence that extends beyond a single performance score.

3.3 Cultural Embeddedness Explained How Meaningful Learning Emerged

MDAF was designed on the assumption that the effectiveness of assessment is determined not only by the system's ability to generate accurate diagnoses but also by the extent to which students perceive assessment tasks as meaningful learning experiences. The analysis in this section evaluates the role of Cultural Embeddedness Perception as a mediating mechanism linking the MDAF to symbolic mathematical performance and student engagement. Table 4 summarizes the test results for this mechanism.

Table 4. Cultural Embeddedness as the Mechanism Underlying Meaningful Learning

Mechanism	Outcome	Standardized Effect (β)	BootSE	95% CI	p	Interpretation
Indirect effect	Symbolic Mathematical Performance	0.18	0.05	[0.08, 0.28]	< .01	Partial mediation
Indirect effect	Student Engagement	0.21	0.05	[0.11, 0.31]	< .01	Partial mediation

The mediation results indicate that Cultural Embeddedness Perception forms part of the pathway through which the MDAF is associated with symbolic mathematical performance and student engagement. The significant indirect effects suggest that students' perceptions of cultural relevance mediate the relationship between the assessment framework and the two learning-related outcomes. Thus, MDAF contributes to meaningful learning not only through its direct relationships with learning outcomes but also through an indirect pathway involving students' perceptions of the cultural relevance of the assessment tasks.

These findings suggest that cultural relevance contributes to meaningful learning by providing a contextual pathway through which students engage with the mathematical tasks. When students perceive assessment tasks as authentic, familiar, and meaningful, they may be better positioned to connect the task context with the mathematical relationships they are expected to interpret and represent. In this sense, cultural embeddedness functions not merely as a contextual feature of the assessment but as part of the learning pathway through which students construct and demonstrate mathematical understanding. This interpretation reinforces the pedagogical role of culturally embedded diagnostic assessment. Rather than treating cultural context merely as a way of presenting mathematical problems, MDAF positions cultural relevance as an element that can contribute to meaningful learning evidence. The findings therefore suggest that the value of AI-supported diagnostic assessment depends not only on its capacity to analyse student responses but also on the meaningfulness of the contexts in which those responses are produced.

3.4 Agreement Between AI-generated Diagnosis and Expert Evaluation

MDAF's reliability depends not only on its ability to improve learning outcomes but also on the accuracy of the diagnoses it generates about students' thought processes. In this section, we evaluated the AI system's diagnostic profiles by comparing them with assessment results from mathematics education experts used as an external reference. The degree of agreement provides evidence of the consistency of the AI's interpretation in identifying the characteristics of students' symbolic mathematical abilities. Table 5 summarizes the analysis results.

Table 5. Agreement Between AI-generated Diagnostic Profiles and Expert Evaluation

Diagnostic Measure	Statistic	Value	Interpretation
Overall symbolic performance	ICC	0.81	Strong agreement
Symbol interpretation	Cohen's Kappa	0.78	Substantial agreement

The high level of agreement between the AI's diagnosis and expert evaluations indicates that the system can consistently interpret students' open-ended responses across various aspects of symbolic mathematical reasoning. These findings suggest that the AI's output is not limited to assigning final scores but can also generate diagnostic profiles that align with professional judgment in identifying students' ability characteristics. Empirical evidence of diagnostic accuracy reinforces MDAF's position as an assessment approach that generates reliable learning evidence for instructional decision-making. The AI's ability to produce diagnostic profiles consistent with expert evaluations creates opportunities to use diagnostic assessments more efficiently without compromising the quality of pedagogical interpretation teachers need when designing follow-up learning activities.

3.5 Empirical Validation of the MDAF

The empirical findings from this study suggest that the relationships observed among the components of MDAF are better understood as an integrated pattern involving cultural context, AI-based analysis, and learning evidence. Rather than establishing a causal pathway, the findings provide empirical support for the proposed structural relationships and the mediating role of Cultural Embeddedness Perception. Figure 5 synthesizes the results, showing how the framework's components are empirically connected.

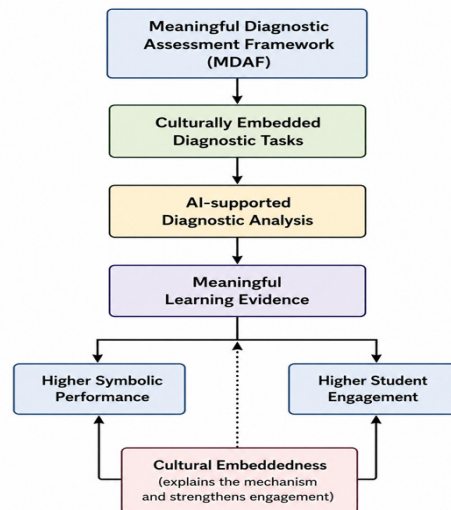


Figure 5. Empirical Validation of the MDAF

Figure 3 shows that diagnostic tasks developed using cultural contexts serve as a starting point for creating relevant learning experiences for students. AI-based analysis then transforms students' open-ended responses into meaningful evidence of learning that not only illustrates achievement levels but also reveals the characteristics of the mathematical reasoning underlying each response. The learning evidence generated through AI-based analysis is associated with symbolic mathematical performance and student engagement within the structural model. At the same time, Cultural Embeddedness Perception represents an indirect pathway through which the MDAF is associated with symbolic mathematical performance and student engagement.

This empirical synthesis shows that MDAF's primary contribution lies not in using artificial intelligence as an automated scoring tool, but in its ability to integrate cultural context, AI-based diagnosis, and pedagogical interpretation into meaningful learning evidence. This perspective shows that diagnostic assessment can provide information that supports understanding of students' thought processes while also forming the basis for more adaptive instructional decision-making. These findings reinforce the argument that AI-based diagnostic assessments will have a more meaningful educational impact when developed through culturally relevant learning experiences.

4. Discussion

For several decades, diagnostic assessment has been understood primarily as a mechanism for identifying initial abilities, detecting misconceptions, and grouping students by level of conceptual mastery. This perspective treats diagnosis as the product of assessment, so the information generated often stops at performance classification. This study's findings indicate that diagnostic assessment should shift from measuring learning to generating meaningful evidence of learning. This shift expands the concepts of "assessment for learning" and "assessment as learning," which view assessment as integral to the learning process (Graven & Venkat, 2021; Zhang, Zhang, et al., 2025). MDAF demonstrates that diagnosis not only identifies learning weaknesses but also generates evidence regarding how students construct symbolic reasoning, select problem-solving strategies, and overcome conceptual obstacles. The main contribution of this study lies in redefining the function of diagnostic assessment as a system for generating meaningful evidence of learning to support more adaptive pedagogical decisions.

Subsequent findings show that culture is not merely the context in which tasks are presented, but rather a mechanism that determines the quality of the learning evidence generated. The mediating relationship suggests that Cultural Embeddedness Perception is associated with the pathway through which students engage with culturally relevant tasks and connect contextual information with more abstract mathematical representations.. These results expand the concept of culturally responsive assessment, which has thus far been more closely associated with equity and inclusivity in learning (Biesta, 2023; Cosme et al., 2025; Oladejo et al., 2023; Prahmana, 2022). These findings are also consistent with situated cognition theory, which posits that knowledge develops through interactions between individuals and their learning contexts (Schlimm, 2025; Xu et al., 2023). In this study, culture serves as an epistemic bridge, enabling the diagnostic process to capture the development of mathematical reasoning more comprehensively.

The mediation results further highlight the role of cultural context in meaningful diagnostic assessment. Cultural Embeddedness Perception partially mediated the relationships between the MDAF and both symbolic mathematical performance and student engagement. This finding suggests that culturally relevant assessment contributes partly

through students' perceptions of the authenticity, familiarity, and meaningfulness of the task context. Rather than indicating that cultural embeddedness changes the strength of the MDAF–outcome relationships, the findings suggest that cultural perceptions form part of the pathway through which the assessment framework is associated with learning-related outcomes (Aldaleel, 2025; McNamara et al., 2023).

Another finding indicates that the primary value of artificial intelligence in assessment lies not in automating scoring but in its ability to interpret evidence of learning from a pedagogical perspective. High agreement between AI diagnoses and expert evaluations suggests that integrating NLP, construct-aligned scoring, and reasoning-pattern analysis can yield diagnoses that are both consistent and meaningful. These results expand the direction of AI research in assessment, which has traditionally focused more on automated scoring and evaluation efficiency (Tan et al., 2023; Zhai et al., 2021). AI is shifting from an automated scorer to a pedagogical interpreter that helps teachers understand students' thought processes. This perspective aligns with developments in learning analytics and explainable artificial intelligence, which emphasise the importance of interpretability in using AI in education (Drijvers & Sinclair, 2024; Payadnya et al., 2025; Weng et al., 2023).

However, the transferability of the MDAF should be considered with caution. The present study examined the framework in a specific mathematics learning context involving junior high school students and culturally embedded diagnostic tasks in Indonesia. The principles underlying the framework—namely, the use of contextually meaningful open-ended tasks, AI-supported analysis of student responses, and pedagogical interpretation of diagnostic evidence—may be applicable to other areas of mathematics and potentially to other subject domains in which students' reasoning can be expressed through interpretable responses (Mahamad et al., 2025). However, the present study does not provide sufficient evidence that the framework would operate the same way across different mathematical topics, educational levels, or cultural settings. Cultural embeddedness may require adaptation to local practices, artefacts, meanings, and students' familiarity with the context (Cosme et al., 2025; Prahmana, 2022). Therefore, studies should examine the broader applicability of MDAF across different mathematical domains, educational populations, and cultural contexts.

Collectively, the findings provide empirical support for the view that meaningful diagnostic assessment can be conceptualised through the integration of cultural context, AI-based analysis, and pedagogical interpretation of students' responses. MDAF positions diagnosis not simply as an endpoint for classifying performance but as a source of evidence that can inform understanding of students' mathematical reasoning and subsequent pedagogical decisions. This study therefore contributes primarily by proposing and empirically examining an integrated assessment framework, rather than establishing a universal causal model of meaningful learning. The findings provide a basis for further investigation of how such an approach may be adapted across mathematical topics, educational settings, and cultural contexts.

5. Conclusion

This study examined the Meaningful Diagnostic Assessment Framework (MDAF) as an approach to integrating culturally embedded diagnostic tasks, AI-supported analysis of open-ended responses, and pedagogical interpretation to generate meaningful evidence of learning. The findings provide empirical support for positive associations between the MDAF and symbolic mathematical performance and student engagement within the structural model. Cultural Embeddedness Perception also partially mediated these relationships, indicating that students' perceptions of the cultural relevance of assessment tasks form part of the pathway linking the framework to the learning-related outcomes. In addition, the relatively high agreement between GradeLab-generated diagnostic profiles and expert evaluations provides evidence supporting the consistency of the AI-supported diagnostic interpretation.

These findings should be interpreted in relation to the design and scope of the present study. Because the study used existing classes and the reported structural analysis does not directly examine between-group changes from baseline to post-intervention outcomes, the findings should not be interpreted as definitive evidence that MDAF causes improvements in mathematical performance or student engagement. Rather, the study provides empirical evidence for the proposed relationships among culturally embedded assessment, AI-supported diagnosis, cultural embeddedness perceptions, and learning-related outcomes.

The transferability of MDAF also requires further investigation. The present study focused on junior high school mathematics within a specific Indonesian cultural context. The core principles of MDAF may be adaptable to other mathematical topics and educational settings, particularly where open-ended responses can provide interpretable evidence of students' reasoning. However, culturally embedded tasks cannot necessarily be transferred directly across contexts because their relevance depends on students' familiarity with local practices, artefacts, meanings, and experiences. Future studies should therefore examine MDAF across different mathematical domains, educational levels, and cultural settings to determine which elements of the framework remain stable and which require contextual adaptation.

Overall, this study demonstrates the feasibility of connecting culturally meaningful diagnostic tasks, AI-supported response analysis, and pedagogical interpretation within an integrated assessment framework. Further experimental and cross-context research is needed to determine the extent to which this approach can improve learning outcomes and to examine its applicability across diverse educational and cultural environments.

Acknowledgments

The authors gratefully acknowledge the financial support provided by Direktorat Penelitian dan Pengabdian kepada Masyarakat, Direktorat Jenderal Riset dan Pengembangan, Kementerian Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia, through the BIMA Research Funding Program 2026. The authors also thank all participants and supporting parties for their valuable contributions to this study.

Author contributions

BA led the study conceptualization, research design, data collection, data analysis, and manuscript drafting. Prof. S contributed to the study design, methodological review, and critical revision of the manuscript. EA contributed to data interpretation, validation of the findings, and critical revision of the manuscript. All authors reviewed and approved the final version of the manuscript. No special authorship arrangements or equal-contribution agreements apply.

Funding

This research was supported by the Direktorat Penelitian dan Pengabdian kepada Masyarakat, Direktorat Jenderal Riset dan Pengembangan, Kementerian Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia, through the BIMA Research Funding Program 2026.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this study.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Redfame Publishing.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

References

- Alam, A., & Mohanty, A. (2023). Cultural beliefs and equity in educational institutions: exploring the social and philosophical notions of ability groupings in teaching and learning of mathematics. *International Journal of Adolescence and Youth*, 28(1). <https://doi.org/10.1080/02673843.2023.2270662>
- Aldaleel, A. M. (2025). Assessing Middle School Teachers' Digital Innovation Skills in Bisha Province: A Comprehensive Study. *Journal of Educational Technology Development and Exchange*, 18(2), 47–66. <https://doi.org/10.18785/jetde.1802.03>
- Alfageh, D. H., York, C. S., Hodge-Zickerman, A., & Xie, Y. (2024). Elementary teachers' use of adaptive diagnostic assessment to improve mathematics teaching and learning: A case study. *International Electronic Journal of*

Mathematics Education, 19(1), em0768. <https://doi.org/10.29333/iejme/14190>

- Andrews, D., Van Lieshout, E., & Bhatta Kaudal, B. (2023). How, Where, And When Do Students Experience Meaningful Learning? *International Journal of Innovation in Science and Mathematics Education*, 31(3). <https://doi.org/10.30722/IJISME.31.03.003>
- Ardianto, M. I., Pamungkas, M. D., & Agustyaningrum, N. (2025). How Does Computational Thinking Affect Mathematical Complex Problem-solving? *Journal of Education and Training Studies*, 14(1), 196. <https://doi.org/10.11114/jets.v14i1.8130>
- Baker, R. S., Gašević, D., & Karumbaiah, S. (2021). Four paradigms in learning analytics: Why paradigm convergence matters. *Computers and Education: Artificial Intelligence*, 2, 100021. <https://doi.org/10.1016/j.caeai.2021.100021>
- Bayaga, A. (2025). Examining Acceptance and Use of AI-powered Adaptive Learning Systems Among Higher Education Instructors—Bridging Digital Gap. *Africa Education Review*, 1–29. <https://doi.org/10.1080/18146627.2025.2578358>
- Bento Miguens, A. L., Nunes Piedade, J. M., dos Santos, R. J. B., & Oliva, T. L. (2024). Meaningful learning in mathematics: a study on motivation for learning and development of computational thinking using educational robotics. *Educational Media International*, 61(1–2), 4–15. <https://doi.org/10.1080/09523987.2024.2357472>
- Biesta, G. (2023). Becoming contemporaneous: intercultural communication pedagogy beyond culture and without ethics. *Pedagogy, Culture & Society*, 31(2), 237–251. <https://doi.org/10.1080/14681366.2022.2164341>
- Chang, W.-C., & Viesca, K. M. (2022). Preparing Teachers for Culturally Responsive/Relevant Pedagogy (CRP): A Critical Review of Research. *Teachers College Record: The Voice of Scholarship in Education*, 124(2), 197–224. <https://doi.org/10.1177/01614681221086676>
- Chen, X., Fu, M., & Li, H. (2026). Enhancing Higher Education through AI-Driven Personalized Adaptive Learning: Evidence from a ChatGPT-Based Strategy. *International Journal of Instruction*, 19(1), 579–594. <https://doi.org/10.29333/iji.2026.19129a>
- Confrey, J., & Shah, M. (2021). Examining instructional change at scale using data from diagnostic assessments built on learning trajectories. *ZDM – Mathematics Education*, 53(6), 1265–1283. <https://doi.org/10.1007/s11858-021-01259-w>
- Cosme, D., Chan, H.-Y., Sinclair, A. H., Benitez, C., Lydic, K., Martin, R. E., ... & Scholz, C. (2025). Perceived self and social relevance of content motivates news sharing across cultures and topics. *PNAS Nexus*, 4(2). <https://doi.org/10.1093/pnasnexus/pgaf019>
- Drijvers, P., & Sinclair, N. (2024). The role of digital technologies in mathematics education: purposes and perspectives. *ZDM – Mathematics Education*, 56(2), 239–248. <https://doi.org/10.1007/s11858-023-01535-x>
- Eglash, R. (2023). *Ethno-biomathematics: A Decolonial Approach to Mathematics at the Intersection of Human and Nonhuman Design* (pp. 289–303). https://doi.org/10.1007/978-3-031-31293-9_18
- Engelhard, Jr., G., & Wang, J. (2024). *Invariant Measurement*. Routledge. <https://doi.org/10.4324/9781003458746>
- Feng, Y., & Hancock, G. R. (2024). A structural equation modeling approach for modeling variability as a latent variable. *Psychological Methods*, 29(2), 262–286. <https://doi.org/10.1037/met0000477>
- Gilmore, C. (2023). Understanding the complexities of mathematical cognition: A multi-level framework. *Quarterly Journal of Experimental Psychology*, 76(9), 1953–1972. <https://doi.org/10.1177/17470218231175325>
- Graven, M., & Venkat, H. (2021). Piloting national diagnostic assessment for strategic calculation. *Mathematics Education Research Journal*, 33(1), 23–42. <https://doi.org/10.1007/s13394-019-00291-0>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). *An Introduction to Structural Equation Modeling* (pp. 1–29). https://doi.org/10.1007/978-3-030-80519-7_1
- Hair, J., & Alamer, A. (2022). Partial Least Squares Structural Equation Modeling (PLS-SEM) in second language and education research: Guidelines using an applied example. *Research Methods in Applied Linguistics*, 1(3), 100027. <https://doi.org/10.1016/j.rmal.2022.100027>
- Han, J., Liu, G., & Xiang, S. (2025). To engage with AI or not: learning engagement among rural junior high school students in an AI-powered adaptive learning environment. *Humanities and Social Sciences Communications*, 12(1), 1292. <https://doi.org/10.1057/s41599-025-05676-0>
- Huang, Y., Zhai, R., & Jiang, L. (2024). A Study on the Impact of Representational Forms of Concepts in Elementary School Mathematics on the Degree of Transfer Occurrence. *Journal of Educational Technology Development and*

- Exchange*, 17(1), 155–171. <https://doi.org/10.18785/jetde.1701.08>
- Kohen, Z., & Nitzan-Tamar, O. (2022). Contextual Mathematical Modelling: Problem-Solving Characterization and Feasibility. *Education Sciences*, 12(7), 454. <https://doi.org/10.3390/educsci12070454>
- Kong, S.-C., & Wang, Y.-Q. (2024). Dynamic interplays between self-regulated learning and computational thinking in primary school students through animations and worksheets. *Computers & Education*, 220, 105126. <https://doi.org/10.1016/j.compedu.2024.105126>
- Kurniawan, I., Sugiman, S., Hamdi, S., & Erlangga, F. (2026). Understanding TPACK Implementation in STEM Mathematics Education: A Phenomenological Study of Indonesian Senior High School Teachers. *Journal of Education and Training Studies*, 14(4), 210. <https://doi.org/10.11114/jets.v14i4.8972>
- Liu, Y. C., Wang, W.-T., & Lee, T.-L. (2021). An Integrated View of Information Feedback, Game Quality, and Autonomous Motivation for Evaluating Game-Based Learning Effectiveness. *Journal of Educational Computing Research*, 59(1), 3–40. <https://doi.org/10.1177/0735633120952044>
- Löfström, E., Koponen, M., Salonen, V., & Aksela, M. (2026). Meaningful Learning in Finnish Mathematics Teaching Online In-Service Training: Two Denotations. *International Journal of Science and Mathematics Education*, 24(5), 57. <https://doi.org/10.1007/s10763-026-10680-0>
- Luo, Q. Z. (2023). The Influence of AI-Powered Adaptive Learning Platforms on Student Performance in Chinese Classrooms. *Journal of Education*, 6(3), 1–12. <https://doi.org/10.53819/81018102t4181>
- Mabalane Jr., D. K. (2025). Analyzing Teachers' Perceptions of Mobile Learning in Senior Secondary Schools: A FRAME Model Perspective in Sub-Saharan Africa. *Journal of Educational Technology Development and Exchange*, 18(2), 127–150. <https://doi.org/10.18785/jetde.1802.07>
- Mahamad, S., Chin, Y. H., Zulmuksah, N. I. N., Haque, M. M., Shaheen, M., & Nisar, K. (2025). Technical Review: Architecting an AI-Driven Decision Support System for Enhanced Online Learning and Assessment. *Future Internet*, 17(9), 383. <https://doi.org/10.3390/fi17090383>
- McNamara, D. S., Arner, T., Butterfuss, R., Fang, Y., Watanabe, M., Newton, N., McCarthy, K. S., Allen, L. K., & Roscoe, R. D. (2023). iSTART: Adaptive Comprehension Strategy Training and Stealth Literacy Assessment. *International Journal of Human–Computer Interaction*, 39(11), 2239–2252. <https://doi.org/10.1080/10447318.2022.2114143>
- Meyer, J., Jansen, T., Daumiller, M., & Fleckenstein, J. (2025). Understanding individual differences in students' responses to technology-based feedback on a writing task: the role of achievement motives and initial task performance. *Journal of Research on Technology in Education*, 1–31. <https://doi.org/10.1080/15391523.2025.2471765>
- Oladejo, A. I., Olateju, T. T., Okebukola, P. A., Agboluaje, T. M., Sanni, R., Shabani, J., ... & Ebisin, A. (2023). The convergence of culture, technology and context: A pathway to reducing Mathophobia and improving achievement in mathematics. *School Science and Mathematics*, 123(2), 82–96. <https://doi.org/10.1111/ssm.12573>
- Payadnya, I. P. A. A., Putri, G. A. M. A., Suwija, I. K., Saelee, S., & Jayantika, I. G. A. N. T. (2025). Cultural integration in AI-enhanced mathematics education: insights from Southeast Asian educators. *Journal for Multicultural Education*, 19(1), 58–72. <https://doi.org/10.1108/JME-09-2024-0119>
- Prahmana, R. C. I. (2022). Ethno-Realistic Mathematics Education: The promising learning approach in the city of culture. *SN Social Sciences*, 2(12), 257. <https://doi.org/10.1007/s43545-022-00571-w>
- Putra, Z. H., Alim, J. A., Kurnia, R., Gunawan, Y., Dahnilsyah, D., & Aljarrah, A. (2026). A systematic literature review of joyful learning in mathematics: an implication for teaching and learning in elementary school. *Frontiers in Education*, 11. <https://doi.org/10.3389/educ.2026.1561649>
- Richardo, R., Wijaya, A., Rochmadi, T., Abdullah, A. A., Nurkhamid, Astuti, A. W., & Hidayah, K. N. (2023). Ethnomathematics Augmented Reality: Android-Based Learning Multimedia to Improve Creative Thinking Skills on Geometry. *International Journal of Information and Education Technology*, 13(4), 731–737. <https://doi.org/10.18178/ijiet.2023.13.4.1860>
- Sacristán, A. I. (2024). Digital Technologies, Cultures and Mathematics Education. *Proceedings of the 14th International Congress on Mathematical Education*, 521–539. https://doi.org/10.1142/9789811287183_0035
- Schlimm, D. (2025). Where Mathematical Symbols Come From. *Topics in Cognitive Science*. <https://doi.org/10.1111/tops.12786>

- Setiyani, Waluya, S. B., Sukestiyarno, Y. L., & Cahyono, A. N. (2022). E-Module Design Using Kvisoft Flipbook Application Based on Mathematics Creative Thinking Ability for Junior High Schools. *International Journal of Interactive Mobile Technologies (IJIM)*, 16(04), 116–136. <https://doi.org/10.3991/ijim.v16i04.25329>
- Shishigu, A., Michael, K., & Atnafu, M. (2026). Effect of blended learning on mathematical achievement and anxiety: A context-based technology integration for meaningful learning. *E-Learning and Digital Media*, 23(4), 324–340. <https://doi.org/10.1177/20427530241241767>
- Sunzuma, G., & Umbara, U. (2025). Ethnomathematics-based technology in Indonesia: A systematic review. *Asian Journal for Mathematics Education*, 4(1), 129–153. <https://doi.org/10.1177/27527263241305812>
- Tan, D., Diatta-Holgate, H. A., & Levesque-Bristol, C. (2023). Perceived autonomy supportive and culturally responsive environments contribute to international students' participation and willingness to communicate. *Current Psychology*, 42(9), 7629–7648. <https://doi.org/10.1007/s12144-021-02063-1>
- Tanase, M., & Kayaalp, D. (2024). From Theory to Practice: Dos of Culturally Responsive Teaching in Teacher Education Programs. *Journal of Education*, 204(4), 790–802. <https://doi.org/10.1177/00220574231221164>
- Ünal, Z. E., Ala, A. M., Kartal, G., Özel, S., & Geary, D. C. (2023). Visual and symbolic representations as components of algebraic reasoning. *Journal of Numerical Cognition*, 9(2), 327–345. <https://doi.org/10.5964/jnc.11151>
- Weng, X., Ng, O.-L., Cui, Z., & Leung, S. (2023). Creativity Development With Problem-Based Digital Making and Block-Based Programming for Science, Technology, Engineering, Arts, and Mathematics Learning in Middle School Contexts. *Journal of Educational Computing Research*, 61(2), 304–328. <https://doi.org/10.1177/07356331221115661>
- Wu, J., Wang, Y., Chen, F., Yin, X., & He, Y. (2025). Impact of AI-Powered Adaptive Learning Platforms on English Reading Proficiency: Evidence From Structural Equation Modeling. *IEEE Access*, 13, 88230–88242. <https://doi.org/10.1109/ACCESS.2025.3571055>
- Wu, X., Zhang, Y., Wu, R., & Chang, H.-H. (2022). A comparative study on cognitive diagnostic assessment of mathematical key competencies and learning trajectories. *Current Psychology*, 41(11), 7854–7866. <https://doi.org/10.1007/s12144-020-01230-0>
- Xu, T., Wu, X., Sun, S., & Kong, Q. (2023). Cognitive diagnostic analysis of students' mathematical competency based on the DINA model. *Psychology in the Schools*, 60(9), 3135–3150. <https://doi.org/10.1002/pits.22916>
- Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, 2021(1). <https://doi.org/10.1155/2021/8812542>
- Zhan, Z., Tu, K., Yang, H., Gao, S., & Zou, X. (2024). Effect of Reverse Engineering Pedagogy on Students' Learning Performance, Cultural Heritage Literacy, and Creativity in a Paper Sculpture C-STEAM Course. *Journal of Educational Technology Development and Exchange*, 17(2), 153–174. <https://doi.org/10.18785/jetde.1702.07>
- Zhang, Y., Zhang, L., Zhang, H., & Wu, X. (2025). Development of a Computerized Adaptive Assessment and Learning System for Mathematical Ability Based on Cognitive Diagnosis. *Journal of Intelligence*, 13(9), 114. <https://doi.org/10.3390/jintelligence13090114>
- Zhang, Y., Zhao, B., Jian, M., & Wu, X. (2025). Cognitive diagnostic analysis of mathematics key competencies based on PISA data. *PLOS ONE*, 20(2), e0315539. <https://doi.org/10.1371/journal.pone.0315539>