International Electronic Journal of Mathematics Education
2026, 21(4), em0905
e-ISSN: 1306-3030
https://www.iejme.com

MODESTUM

Research Article

OPEN ACCESS

A digital two-tier assessment model for mathematical creative problem-solving: Integrating ethnomathematics and second-order CFA

Himmatul Ulya [1]*, Ratri Rahayu [1], Alif Catur Murti [1], Aivin B. Balquedra [2]

[1] Universitas Muria Kudus, INDONESIA
[2] Nueva Ecija University of Science and Technology, PHILIPPINES
***Corresponding Author:** himmatul.ulya@umk.ac.id**

Citation: Ulya, H., Rahayu, R., Murti, A. C., & Balquedra, A. B. (2026). A digital two-tier assessment model for mathematical creative problem-solving: Integrating ethnomathematics and second-order CFA. *International Electronic Journal of Mathematics Education, 21*(4), em0905. https://doi.org/10.29333/iejme/19964

ARTICLE INFO

Received: 15 May 2026

Accepted: 24 Sep. 2026

ABSTRACT

Conventional mathematics assessments often emphasize final answers and provide limited evidence of students' mathematical creative problem-solving (MCPS) processes. This study developed and evaluated an ethnomathematics-based digital two-tier assessment that combines complex multiple-choice items in Tier 1 and open-ended reasoning tasks in Tier 2. This R&D administered the instrument to 240 eighth-graders from two Indonesian junior high schools. The researchers examined the construct validity through second-order confirmatory factor analysis, while assessing the reliability using construct reliability. The global fit test results suggest that the model has satisfactory global fit (χ^2 = 22.36, df = 16, p = .132, RMSEA = .041, GFI = .98, AGFI = .95, NNFI = .98). The examination of overall construct reliability also yields a score of .82. Therefore, based on the results, the researchers considered the instrument as a reliable composite measure of MCPS. However, the strategic flexibility dimension requires further refinement before the instrument can be interpreted as four independent subscales.

Keywords: CFA, digital assessment, ethnomathematics, mathematical creative problem-solving, two-tier assessment

INTRODUCTION

Mathematics education in the 21st century requires students to reproduce formulas, follow familiar procedures, interpret unfamiliar situations, generate alternative strategies, justify their reasoning, and adapt prior knowledge to new problems. Therefore, the development of those competencies requires mathematical creative problem-solving, MCPS. This problem-solving includes formulating, exploring, evaluating, and refining solution skills to non-routine mathematical tasks (Leikin & Pitta-Pantazi, 2013; Silver, 1997; van Hooijdonk et al., 2020). MCPS therefore depends not only on conceptual and procedural knowledge but also on students' skills to think flexibly and evaluate the accuracy of various approaches (Bicer, 2021; Hurrell, 2021; Schoevers et al., 2020).

In terms of mathematical creativity, students require five capabilities such as fluency, flexibility, originality, and elaboration (Antika et al., 2023; Hamid & Kamarudin, 2020; Kattou et al., 2013). On the other hand, mathematics also has two operational dimensions within valid mathematical reasoning. Firstly, divergent thinking, this dimension supports the generation of multiple ideas, representations, and strategies. Secondly, convergent thinking, this dimension enables students to examine, compare, and select mathematically appropriate solutions (Leikin & Pitta-Pantazi, 2013; Meier et al., 2021; Sipahi & Bahar, 2025). Thus, assessing MCPS requires attention to both the solutions students generate and the reasoning processes of the developed solutions.

Conventional mathematics assessments remain limited in this respect due to the emphases on correctness, procedural fluency, and standard algorithms. Despite the excellence of such assessment to determine the students' expected outcomes, the assessment could only reveal the mechanism of problem interpretation by the students, strategical shift, unsuccessful revision attempt, or conclusion justification (Asamoah, 2019; Boaler et al., 2022; Hirsh, 2020). For a creativity assessment, this limitation facilitates to understand the incapability of solely final answer to generalize strategic flexibility, originality, and elaboration (Bicer et al., 2024; Keleş, 2022; Lee et al., 2003). Recent reviews also indicate the common measurement of mathematical creativity through product-oriented open-ended tasks. Therefore, students' reasoning processes and the interaction between divergent and convergent thinking receive less systematic attention (Sipahi & Bahar, 2025; Suherman & Vidákovich, 2022).

Copyright © 2026 by Author/s and Licensed by Modestum Open DOO, Serbia. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The implementation of two-tier assessment could overcome the limitations due to the combined selected-response and constructed responses format. The current model has tier 1 assessment by using complex multiple-choice items. These items are useful to examine students' interpretation and selection of plausible mathematical solutions. On the other hand, the applied tier 2 assessment demands open-ended explanations dealing with the realized reasoning, representations, or alternative strategies underlying those selections. This structure could provide accurate solution and qualified reasoning to examine within the same task. This model is also relevant to MCPS since creative reasoning is not always associated with correct answer. In fact, an incomplete answer could provide a productive and original solution process.

The quality of MCPS assessment also relies on the nature of the task context. The solution upon routine exercises may only require direct recall, whereas non-routine problems require students to identify unfamiliar association, reorganize prior knowledge, and determine relevant mathematical ideas (Faulkner et al., 2023; Hurrell, 2021; van Hooijdonk et al., 2020). Ethnomathematics provides a meaningful basis for constructing such tasks by situating mathematics within cultural practices, artefacts, and community activities (D'Ambrosio, 1985; Orey & Rosa, 2007; Rosa et al., 2016). Integrating cultural elements into mathematics curricula can make mathematical learning more meaningful and contribute to students' motivation and achievement (Fouze & Amit, 2018; Mei et al., 2025; Sutarto et al., 2022; Ulya & Rahayu, 2021).

Routine cultural practices may reveal intradisciplinary, interdisciplinary, and sociocultural connections between community-based mathematical activities and institutional mathematics as the realized extensions of the perspectives (Kaya & Yavuz, 2025; Mei et al., 2025; Rodríguez-Nieto & Alsina, 2022). In the Indonesian context, most ethnomathematics research only focused on geometry and cultural forms such as traditional architecture, batik, handicrafts, and local practices. Therefore, the substantial mathematical potential of culturally situated contexts is important (Azmi et al., 2025; Wibawa et al., 2025). More recently, an ethnomathematical didactic sequence grounded in chontaduro commercialization enabled primary school students to engage in additive and multiplicative problem-solving. These engagement require practices embedded in daily sociocultural environment (Rodríguez-Nieto et al., 2026). These studies reinforce the importance of cultural context as mathematical structure to contribute directly in solving problems, mathematizing, comparing strategies, and justifying solutions instead of being decorative narratives.

Despite growing interest in ethnomathematics, assessment-oriented research remains relatively limited. Indonesian ethnomathematics research published between 2016 and 2022 mostly applied qualitative studies, whereas quantitative and development-oriented approaches were less prevalent (Turmuzi et al., 2023). A more recent review further demonstrates the breadth of cultural-mathematical exploration in Indonesia while identifying the need for further investigation of mathematical concepts across diverse regional cultures (Wibawa et al., 2025). At the classroom level, practical barriers to ethnomathematics integration include limited teacher knowledge of ethnomathematics approaches, drawbacks in geometry content knowledge, and resistance to established changing teaching practices (Sunzuma & Maharaj, 2019). Therefore, those findings suggest the importance explicit, usable, and empirically validated tools for translating ethnomathematical ideas into classroom assessment instead of merely broader cultural mapping and instructional designs. In this case, mathematical creativity crucially needs the tools especially those with excellent validity and reliability to capture multidimensional reasoning (Sipahi & Bahar, 2025; Suherman & Vidákovich, 2022).

Ethnomathematical tasks have excellent relevance to transfer knowledge for the students. This situation may occur when students apply the previous learned mathematical problems to the latest given problems with similar structure to the original tasks. This type of transfer is familiar with the term of near transfer. On the other hand, the 'far transfer' requires the students to apply their knowledge to different-substantial contexts in representation, setting, or problem demands (Barnett & Ceci, 2002; Orón & Lizasoain, 2023). In the present assessment, the researchers designed the cultural tasks to vary in terms of familiarity and structural complexity. Therefore, students must apply the previously learned concepts to the closely related situations. The students could also reorganize some ideas into less-familiar contexts. The two-tier format facilitates the examining process of the students' appropriate answer selection and students' prior-knowledge connection to the demands of the problems.

Digital technology further supports the administration and documentation of this assessment model. Digital assessment can standardize item presentation, automate scoring, store open-ended responses, and organize data efficiently for further analysis (Holmes et al., 2019; Susanti & Yuhendri, 2024; Ulya et al., 2024b). Nevertheless, digitalization is meaningful only when it contributes to the assessment process. In this study, the platform integrates both tiers, records and scores Tier 1 responses, stores Tier 2 explanations, and organizes student data for rubric-based scoring and psychometric analysis. The technology therefore functions not merely as a delivery medium but as an infrastructure for linking response accuracy with reasoning evidence.

The present study extends the previous works by Ulya et al. (2024a), the MCPS instrument design for JHS students with the expert judgment as the content validator. Those previous authors developed the initial conceptual framework, incorporated real-world problems, SOLO taxonomy levels, and near- and far-transfer tasks, and provided evidence that the items were relevant to the intended construct. The current study extends that work by converting the instrument into a digital two-tier format, embedding more explicit ethnomathematical contexts, and examining its empirical structure through second-order confirmatory factor analysis. It also evaluates whether numerical fluency, strategic flexibility, structural innovation, and analytical integration can be represented as related first-order dimensions of a higher-order MCPS construct.

Recent studies therefore demonstrate increasing sophistication in the use of ethnomathematics for curriculum design, the identification of mathematical connections in cultural practices, and contextualized problem-solving instruction (Fouze & Amit, 2018; Rodríguez-Nieto et al., 2026; Rodríguez-Nieto & Alsina, 2022). However, these lines of research primarily address curriculum, pedagogy, or the identification and use of mathematical ideas embedded in cultural practices rather than the psychometric validation of assessment instruments. In parallel, mathematical creativity studies often emphasize the products of divergent thinking, while digital assessments frequently prioritize delivery and administrative efficiency (Bicer et al., 2024; Sipahi & Bahar, 2025; Suherman & Vidákovich, 2022; Turmuzi et al., 2023). Consequently, limited evidence is available on an integrated assessment

model that combines authentic ethnomathematical problems, selected- and constructed-response formats, divergent and convergent thinking, near- and far-transfer, digital administration, and hierarchical construct validation.

Accordingly, this study aimed to develop and evaluate an ethnomathematics-based digital two-tier assessment for measuring junior high school students' mathematical creative problem-solving. The study addressed the following research questions: (1) how was the ethnomathematics-based digital two-tier assessment designed to represent mathematical creative problem-solving through divergent and convergent thinking, non-routine cultural problems, and near- and far-transfer?; and (2) to what extent does the proposed second-order measurement model demonstrate acceptable construct validity and reliability for assessing mathematical creative problem-solving?

MATERIALS AND METHODS

Research Design

This study used a research-and-development (R&D) design adapted from the framework introduced by Borg and Gall (1989). The approach was selected to guide the development and validation of an ethnomathematics-based digital two-tier assessment intended to measure mathematical creative problem-solving (MCPS). The procedure was organized into six successive stages: (1) needs analysis, (2) instrument design, (3) expert validation, (4) instrument revision, (5) pilot implementation, and (6) psychometric evaluation. Together, these stages provided a structured basis for ensuring that the assessment was theoretically informed, feasible for classroom use, and supported by empirical evidence.

The researchers designed the instrument by combining mathematical creativity, ethnomathematical context, and digital assessment technology into a two-tier format. In this research, the tier 1 consisted of complex multiple-choice to examine students' conceptual understanding and strategic decisions. On the other hand, the tier-2 assessment required the students to explain, justify, or generate alternative solutions. In this assessment tier, the students shared the responses with open-ended type responses. The researchers expected the assessment design to capture the complementary roles of both convergent and divergent thinking in mathematical creative problem-solving.

Once the researchers completed the development process, the researchers performed psychometric quality test of the instrument with second-order confirmatory factor analysis, CFA, by using Lisrel 8.50. This analysis process evaluated the four first-tier dimensions such as numerical fluency, strategic flexibility, structural innovation, and analytical integration could reflect the higher-order MCPS construct. The researchers also determined the content validity with expert judgment using Aiken's V coefficient (Aiken, 1980). For the construct reliability, the researchers performed standardized loading factors and measurement error values.

Needs Analysis

Prior to the instrument development, the researchers reviewed some literature works and performed a need analysis by involving mathematics teachers at JHS in Kudus Regency, Central Java, Indonesia. After reviewing the literature works, the researchers found consistent indication from the previous studies, suggesting the integration between divergent and convergent thinking in terms of mathematical creative problem-solving skills, MCPS. Nevertheless, mathematics assessments remain predominantly outcome-oriented, emphasizing correct answers. This assessment led to limited opportunities to examine students' reasoning processes (Sipahi & Bahar, 2025; Suherman & Vidákovich, 2022).

The needs analysis revealed classroom assessment challenges. Most teachers reported the conceptual understanding and procedural fluency as the focus through routine textbook-based questions. This situation limited the students' opportunities to demonstrate mathematical creativity. Teachers also indicated difficulties in assessing students' creative reasoning due to the unavailability of standardized instrument to evaluate alternative strategies, justify, and provide non-routine problem-solving matters. In addition, many students experienced difficulties when solving unfamiliar mathematical problems. Therefore, the students tended to rely on memorized procedures instead of developing alternative solution strategies. These findings highlighted the need for an assessment capable of measuring both mathematical performance and the reasoning processes underlying creative problem-solving.

In this research, the researchers analyzed the needs in January 2025, involving seven mathematics teachers of JHS in Kudus Regency. Then, the researchers collected the data with semi-structured interviews to reveal the teachers' current assessment practices, challenges in evaluating mathematical creativity, and necessity upon a process-oriented assessment. The researchers analyzed the data descriptively by identifying recurring themes and common instructional challenges. The results informed the assessment developments starting from the selected two-tier format, the four MCPS dimensions, the integration of ethnomathematical contexts, and the design of near-transfer and far-transfer tasks.

Participants and Setting

Prior to the research, the researchers conducted a pilot study by involving 240 eighth-graders from two public JHS in Kudus Regency, Central Java, Indonesia. The researchers applied multistage cluster sampling to select the schools as the institutional representations of implementing the national mathematics curriculum under comparable learning situations. Then, the researchers determined the student participation purposively with three inclusion criteria. Firstly, the current enrolment should only the eighth graders. Secondly, the classes used an instrument to assess the completion of the mathematical topics. Thirdly, the participants had to willingly participate proven with written consent from the parents or their guardians. In this research, the researchers excluded the students who failed to complete all assessments items from final analysis.

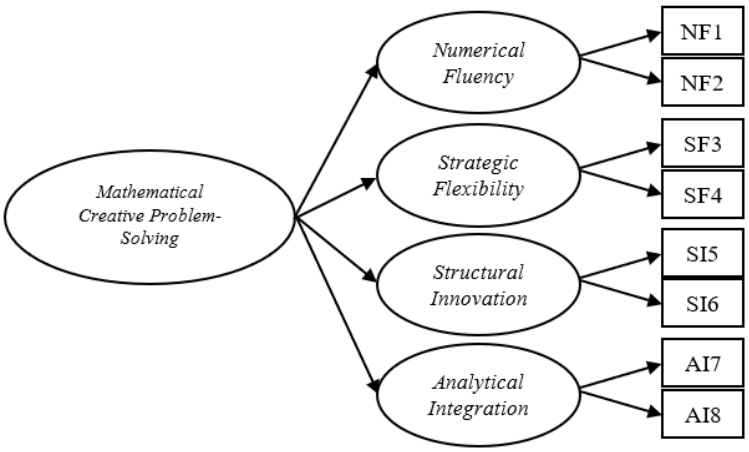


Figure 1. Proposed measurement model of the ethnomathematics-based digital two-tier MCPS assessment (Source: Authors' own elaboration)

The researchers administered the assessment during routine mathematics lesson. In this process, the researchers used a web-based platform with a specific development for the current study. Prior to the assessment administration, the researchers informed the students about the test procedures. The researchers also provided adequate time for the students to understand and getting familiar with the digital interface. This participation was voluntarily and the researchers labelled the responses anonymously to keep the participant confidentiality.

The researchers selected the eighth-graders since they learned the mathematical concepts incorporated into the assessment. The students also already had adequate cognitive readiness to engage with non-routine tasks with mathematical creativity requirement. The students, under the current grade, could also provide an appropriate setting for examining the mechanism of ethnomathematical context interactions with both divergent and convergent thinking in the development of mathematical creative problem-solving.

Instrument Development

The assessment instrument was developed to capture mathematical creative problem-solving (MCPS) as a multidimensional construct comprising four closely connected dimensions: Numerical Fluency (NF), Strategic Flexibility (SF), Structural Innovation (SI), and Analytical Integration (AI). These dimensions were grounded in the mathematical creativity frameworks advanced by Silver (1997) and Leikin & Pitta-Pantazi (2013), and were subsequently refined and operationalized by Ulya et al. (2024a). Taken together, they reflect students' capacity to generate, assess, and synthesize mathematical ideas when responding to non-routine problems.

Figure 1 illustrates the proposed measurement model of the ethnomathematics-based digital two-tier assessment. The model conceptualizes mathematical creative problem-solving (MCPS) as a higher-order construct represented by four first-order dimensions: Numerical Fluency, Strategic Flexibility, Structural Innovation, and Analytical Integration, each measured by two observed indicators.

The final version of the instrument contained eight two-tier items, with two items assigned to each MCPS dimension. Each task consisted of two interdependent parts. Tier 1 presented complex multiple-choice alternatives and was intended to assess conceptual understanding and convergent decision-making. Tier 2, in contrast, asked students to explain, justify, or formulate alternative solution strategies through open-ended responses. By linking answer selection with written reasoning, the instrument was designed to assess both the products of students' mathematical work and the cognitive processes supporting those products. This arrangement provided more comprehensive evidence of creative mathematical thinking than a conventional single-tier format.

The structure of observed learning outcome, the SOLO taxonomy, particularly consisting of the unistructural, multi-structural, and relational levels also informed the item construction. In this research, the researchers used the taxonomy to organize the progression of cognitive demand, started with identifying relevant mathematical elements, applying related elements, and culminating in the element integration into a coherent solution. Therefore, the instrument evaluated both the correct answers, the students' organizations, and the students' mathematical reasonings.

Both near-transfer and far-transfer tasks were incorporated to examine students' skills to apply prior knowledge across different situations (Barnett & Ceci, 2002). Near-transfer items asked students to use familiar mathematical concepts in culturally recognizable situations that retained structural similarities to previously learned tasks. Far-transfer items, by comparison, placed students in less familiar ethnomathematical situations that required them to restructure prior knowledge, identify new relationships, and devise alternative solution paths. The combination of these task types enabled the instrument to distinguish routine procedural application from more flexible and adaptive mathematical reasoning.

The ethnomathematical content was drawn from cultural practices in Central Java, including traditional architecture, batik motifs, local trading activities, and cultural artefacts. These contexts were selected because they naturally embody mathematical ideas such as geometry, ratio and proportion, measurement, symmetry, and pattern recognition. The cultural elements were not

Table 1. Sample two-tier assessment item representing numerical fluency

Component	Description
MCPS Dimension	Numerical Fluency (NF2)
SOLO Level	Relational
Transfer Type	Near-transfer
Mathematical Topic	Geometry and Measurement
Ethnomathematical Context	Borobudur temple relief panel
Learning Indicator	Determine the length, width, and area of a rectangle with a given perimeter expressed as integers.
Tier 1	An artisan produces a rectangular replica of a Borobudur relief panel with a perimeter of 120 cm. The length is greater than the width, and both dimensions are integers. Which of the following represents the correct dimensions and area? A. 46 cm × 16 cm; 690 cm^2 B. 44 cm × 16 cm; 704 cm^2 C. 37 cm × 23 cm; 851 cm^2 D. 32 cm × 27 cm; 897 cm^2
Tier 2	If you found another possible solution, write two different pairs of dimensions that satisfy the same conditions and explain your reasoning.
Scoring	Tier 1 = 0–2 points; Tier 2 = 0–2 points (maximum = 4 points).

Table 2. Analytic scoring rubric for Tier 2

Score	Criteria
2	Provides two mathematically correct alternative solutions with clear and logical justification.
1	Provides one correct alternative solution or partially correct reasoning.
0	Provides no valid alternative solution or no mathematical justification.

used merely as decorative backgrounds. Instead, they formed part of the mathematical structure of the tasks and required students to interpret authentic situations, identify relevant relationships, and generate creative solutions. In this way, the instrument assessed mathematical creativity while also examining students' skills to connect formal mathematical concepts with culturally meaningful experiences.

Table 1 illustrates the two-tier item applied in the assessment. In this research, tier 1 evaluated the students' identification skills upon appropriate mathematical solutions. On the other hand, the tier 2 elicits the alternative solution strategies and reasoning supports. Therefore, the items have the capability to capture both convergent and divergent thinking within the ethnomathematical context, providing opportunities for the researchers to assess students' creativity comprehensively. Tier 2 responses were evaluated using an analytic scoring rubric ranging from 0 to 2, as presented in **Table 2**. A score of 2 was assigned when students provided two mathematically correct alternative solutions supported by clear and logical justification. A score of 1 indicated one correct alternative solution or partially correct reasoning, whereas a score of 0 reflected the absence of a valid alternative solution or mathematical justification.

Digital Assessment Platform

The researchers used a web-based platform, with specific purpose to the current development, to share to the students. The researchers argued this platform to have the capability of packing all elements of the digital two-tier assessment into a single environment. Thus, the platform could enable students completing tier 1 and tier 2 sequentially without switching from different applications of interfaces.

During the operation, the system displayed the items, randomized the sequence to lower order-related response bias, recorded the required time to answer the items, and scored the tier-1 responses automatically. For tier-2 question items, the system automatically saved the answers to evaluate with an analytic scoring rubric. The system also arranged student responses in a structured database, thereby supporting efficient data management and subsequent psychometric analysis.

The platform was designed to serve a broader function than simply converting a paper-based test into an online format. Its central purpose was to support process-oriented assessment by recording both students' selected answers and the explanations underlying those choices. This dual documentation made it possible to examine mathematical creative problem-solving in greater depth. In addition, the digital system streamlined test administration, minimized manual data-entry errors, and promoted consistent assessment procedures across the two participating schools.

Scoring of the assessment differed across the two tiers. Tier 1 responses were scored automatically by the digital platform based on predetermined answer keys. Tier 2 responses were evaluated by a trained evaluator using the finalized analytic rubric, which specified performance criteria for the four MCPS dimensions. When responses were ambiguous, scoring decisions were made by referring to the rubric descriptors to ensure consistency. Because all open-ended responses were evaluated by a single trained evaluator, inter-rater reliability was not calculated. This limitation is considered in the interpretation of the findings and is discussed further in the Discussion section.

Content Validation

Before the instrument was administered in the field, a content-validation procedure was conducted to examine the relevance, clarity, and representativeness of each item. The expert panel consisted of five specialists, including two experts in mathematics

Table 3. Content validity results based on Aiken's V coefficient

Item No.	Item Code	Aiken's V	Critical Value	Decision
1	NF1	0.85	0.8	Valid
2	SF4	0.9	0.8	Valid
3	NF2	0.95	0.8	Valid
4	SF3	0.85	0.8	Valid
5	SI6	0.85	0.8	Valid
6	AI7	0.9	0.8	Valid
7	SI5	0.95	0.8	Valid
8	AI8	0.85	0.8	Valid

education, two experts in ethnomathematics, and one expert in educational measurement. All experts independently reviewed the eight items and assigned ratings using Aiken's validation procedure.

The degree of agreement among the experts was quantified using Aiken's V coefficient. This coefficient indicates the extent to which an item is judged to represent the intended construct. In accordance with the criterion applied in this study, items with Aiken's V values above 0.80 were regarded as acceptable for empirical testing.

As reported in **Table 3**, the coefficients ranged from 0.85 to 0.95, demonstrating a high level of agreement across the expert panel. All eight items exceeded the minimum criterion, indicating that they appropriately represented the four dimensions of MCPS. The experts nevertheless proposed several minor refinements, particularly concerning the clarity of wording, the suitability of cultural contexts, and the consistency of ethnomathematical representations. These recommendations were incorporated before the pilot implementation. Since every item met the required content-validity standard, none was removed from the instrument.

Taken together, these findings provided preliminary evidence that the instrument was aligned with the conceptual structure of mathematical creative problem-solving and was sufficiently appropriate to proceed to psychometric evaluation.

Ethical Considerations

Ethical clearance for this study was granted by the Research Ethics Committee of Universitas Muria Kudus under Ethical Approval No. 292/LPPM.UMK/B.09.129/VI/2025. In this research, the participation was voluntary. Before collecting the data, formal permission was obtained from the participating schools. Then, the researchers used written informed consent from the students and their parents or guardians. In this recruiting process, the researchers informed all participants, especially the students, about any possible involvement discontinuation when the students found the process inconvenient. In this case, the researchers asserted that no academic consequences should the participants experience.

After collecting the responses, the researchers anonymized them before further analyses. The researchers also did not retain any identifying information in the research database. The researchers solely used the collected data for research purposes and managed them based on the established ethical principals for studies with minors' involvement. The entire procedure followed institutional requirements related to informed consent, confidentiality, voluntary participation, and responsible data handling.

Data Analysis

The psychometric performance of the developed instrument was evaluated through second-order Confirmatory Factor Analysis (CFA) using LISREL 8.50. This procedure was employed to determine whether the four first-order dimensions (Numerical Fluency, Strategic Flexibility, Structural Innovation, and Analytical Integration) appropriately reflected the higher-order construct of mathematical creative problem-solving (MCPS). The adequacy of the measurement model was examined using several goodness-of-fit indices, including the Chi-square statistic, *p*-value, RMSEA, GFI, AGFI, NNFI, CFI, and IFI. A model was considered acceptable when the Chi-square test was non-significant, RMSEA was ≤ 0.08, GFI and AGFI were ≥ 0.90, NNFI was ≥ 0.90, and both CFI and IFI reached values of at least 0.95.

Construct validity at the indicator level was examined using standardized loading factors (SLF) together with their corresponding *t*-values. Indicators with SLF values of 0.30 or higher and absolute *t*-values greater than 1.96 were regarded as statistically acceptable. In addition, convergent validity was assessed through the Average Variance Extracted (AVE), where values of 0.50 or above indicated adequate convergence. The internal consistency of the instrument was evaluated using Construct Reliability (CR), calculated from the standardized factor loadings and measurement error estimates. Following the recommendation of Hair et al. (2019), CR values exceeding 0.70 were interpreted as evidence of satisfactory reliability.

RESULTS

Pilot Implementation

The pilot study was conducted with 240 eighth-graders from two public junior high schools. Participants completed the eight-item digital two-tier assessment within 120 minutes. Responses to Tier 1 were scored automatically by the web-based system, while the open-ended answers submitted in Tier 2 were assessed using the predefined analytic rubric. The resulting scores were then used to evaluate the empirical structure and reliability of the MCPS instrument.

Table 4. Goodness-of-fit indices for the second-order CFA model

Goodness-of-Fit Index	Acceptance Criterion	Obtained Value	Interpretation
Chi-square	< 2df	22.36	Fit
Degrees of freedom	-	16	-
p-value	≥ 0.05	0.132	Fit
RMSEA	≤ 0.08	0.041	Fit
GFI	≥ 0.90	0.98	Fit
AGFI	≥ 0.90	0.95	Fit
NNFI	≥ 0.90	0.98	Fit
CFI	≥ 0.95	0.99	Fit
IFI	≥ 0.95	0.99	Fit
Standardized RMR	≤ 0.08	0.031	Fit

Table 5. Item-level CFA results

Dimension	Item	Loading	t-value	Error Variance	Interpretation
Numerical Fluency	NF1	0.64	fixed	0.59	Acceptable
	NF2	0.67	8.11	0.55	Acceptable
Strategic Flexibility	SF3	0.53	fixed	0.72	Acceptable
	SF4	0.45	5.83	0.80	Relatively weak
Structural	SI5	0.67	fixed	0.55	Acceptable
Innovation	SI6	0.65	7.62	0.58	Acceptable
Analytical	AI7	0.68	Fixed	0.54	Acceptable
Integration	AI8	0.70	8.58	0.51	Acceptable

Note: "Fixed" indicates reference indicators constrained for model identification.

Adequacy of the Data for Factor Analysis

Prior to performing Confirmatory Factor Analysis (CFA), the suitability of the sample and correlation matrix was assessed. The Kaiser–Meyer–Olkin value was 0.882, which was well above the minimum threshold of 0.50. This finding showed that the response data had sufficient intercorrelation and were appropriate for factor analysis. The proposed construct structure was subsequently tested through second-order CFA using LISREL 8.50.

Goodness-of-Fit of the Measurement Model

The hypothesized model positioned Numerical Fluency, Strategic Flexibility, Structural Innovation, and Analytical Integration as four first-order dimensions underlying the higher-order construct of mathematical creative problem-solving. The overall model-fit results are presented in **Table 4**.

The Chi-square statistic was 22.36 with 16 degrees of freedom and a non-significant *p*-value of 0.132. The RMSEA value of 0.041, with a 90% confidence interval from 0.000 to 0.078, also indicated a close fit between the model and the observed data. In addition, the GFI, AGFI, NNFI, CFI, and IFI values all exceeded their recommended cut-off points. The standardized RMR of 0.031 further suggested only a small discrepancy between the observed covariance matrix and the covariance matrix implied by the model. Taken together, these indicators supported the satisfactory global fit of the proposed measurement structure.

Item-Level Construct Validity

Table 5 presents the completely standardized factor loadings for the eight observed indicators. The coefficients ranged from 0.45 to 0.70. Seven indicators achieved loading values of at least 0.50, while SF4 produced a lower loading of 0.45. Even so, all indicators loaded positively on their designated first-order dimensions.

AI8 produced the highest loading at 0.70, followed by AI7 at 0.68. NF2 and SI5 each recorded a loading of 0.67, while SI6 and NF1 yielded values of 0.65 and 0.64, respectively. SF3 loaded at 0.53, whereas SF4 showed the lowest contribution at 0.45. The weaker loading of SF4 indicates that this item may capture Strategic Flexibility less effectively than the other indicators capture their corresponding dimensions. For this reason, SF4 should be reviewed and refined in future validation studies.

The *t*-values for the freely estimated item parameters all exceeded the critical value of 1.96: NF2 = 8.11, SF4 = 5.83, SI6 = 7.62, and AI8 = 8.58. In each dimension, the first indicator was fixed as a reference item for model identification, so LISREL did not provide a separate *t*-value for those parameters. The significant estimates for the freely estimated loadings support the relationships between the observed indicators and their respective first-order constructs. **Figure 2** presents the t-value solution for the second-order CFA model. All freely estimated item loadings exceeded the critical value of 1.96, indicating statistically significant relationships between the observed indicators and their corresponding first-order dimensions.

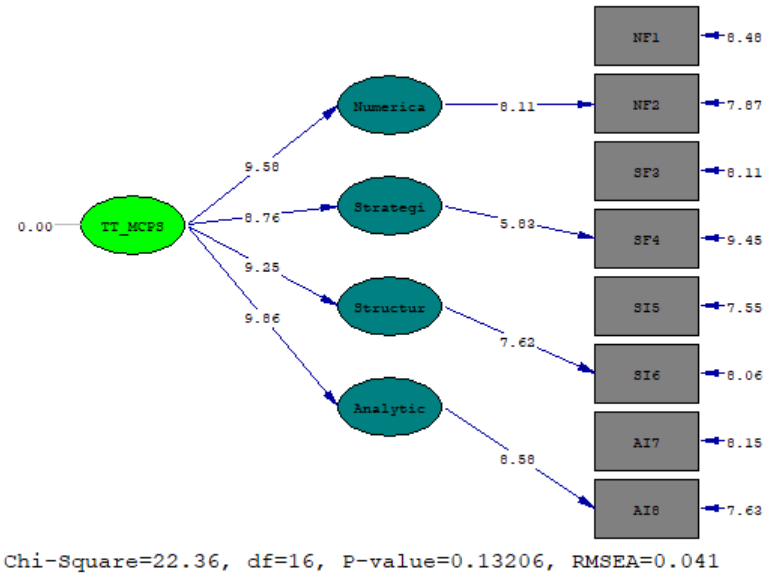


Figure 2. T-value solution of the second-order CFA model (Source: Authors' own elaboration)

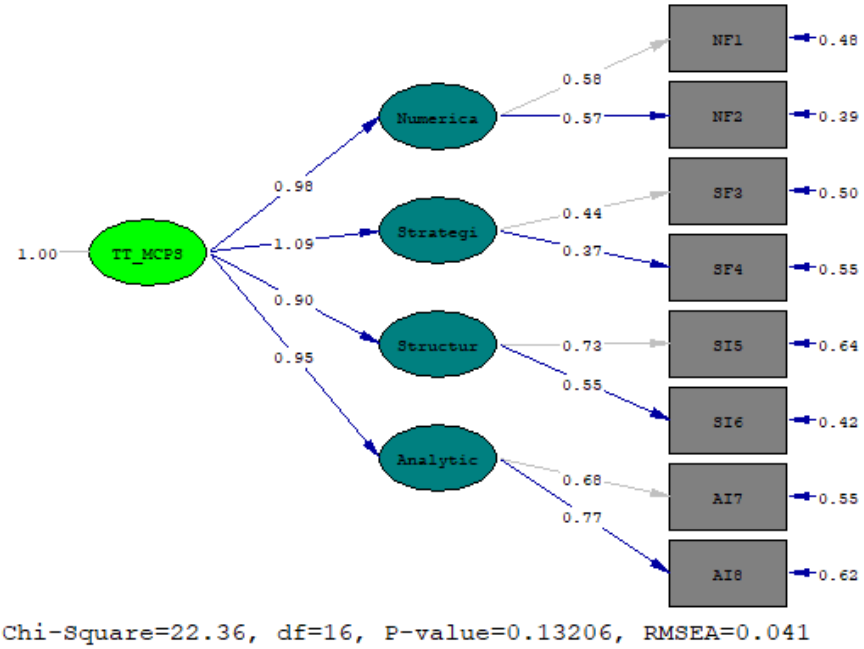


Figure 3. Completely standardized solution of the second-order CFA model (Source: Authors' own elaboration)

Second-Order Structure of MCPS

At the second-order level, Numerical Fluency, Strategic Flexibility, Structural Innovation, and Analytical Integration were modeled as dimensions of the broader MCPS construct. The paths from MCPS to Numerical Fluency, Structural Innovation, and Analytical Integration were positive and statistically significant, with *t*-values of 9.58, 9.25, and 9.86, respectively. The relationship between MCPS and Strategic Flexibility was likewise significant, with a *t*-value of 8.76.

Nevertheless, the standardized second-order loading for Strategic Flexibility was 1.09. This estimate was accompanied by a negative residual variance of -0.18 and a squared multiple correlation greater than 1.00. LISREL therefore issued a warning indicating that the PSI matrix was not positive definite. This result reflects an improper or Heywood-type solution at the second-order level. Accordingly, despite the satisfactory global fit statistics, the second-order estimates, especially those related to Strategic Flexibility, should be interpreted with caution. **Figure 3** displays the completely standardized estimates of the second-order CFA model. The estimate for Strategic Flexibility exceeded 1.00, supporting the need for cautious interpretation and further refinement of this dimension.

As shown in **Figure 3**, the standardized loading for Strategic Flexibility exceeded 1.00, indicating an improper (Heywood-type) estimate that should be interpreted cautiousl

Table 6. Convergent validity and reliability by MCPS dimension

Dimension	AVE	CR	Interpretation
Numerical Fluency	0.43	0.60	Below recommended thresholds
Strategic Flexibility	0.24	0.39	Below recommended thresholds
Structural Innovation	0.44	0.61	Below recommended thresholds
Analytical Integration	0.48	0.64	Approaches the AVE criterion; CR remains below 0.70

Convergent Validity and Construct Reliability

Average Variance Extracted (AVE) and Construct Reliability (CR) were calculated for each first-order dimension using the completely standardized loadings and corresponding measurement-error estimates. The results are provided in **Table 6**.

None of the first-order dimensions met the conventional AVE threshold of 0.50. Analytical Integration obtained the highest AVE value at 0.48, followed by Structural Innovation at 0.44, Numerical Fluency at 0.43, and Strategic Flexibility at 0.24. The dimension-level CR values ranged from 0.39 to 0.64, all of which fell below the commonly recommended minimum of 0.70. These results suggest that convergent validity and internal consistency were not yet fully established at the individual-dimension level, with Strategic Flexibility showing the weakest performance.

Despite these limitations, the overall construct reliability coefficient for the complete eight-item MCPS instrument was 0.82. This value surpassed the recommended criterion of 0.70 and indicated satisfactory internal consistency when the instrument was treated as a composite measure of MCPS. However, the contrast between the satisfactory overall reliability and the lower reliability coefficients at the dimension level suggests that the instrument is currently better supported as a general MCPS measure than as four separate subscales.

Summary of Psychometric Findings

The second-order CFA model demonstrated satisfactory overall fit according to the Chi-square, RMSEA, GFI, AGFI, NNFI, CFI, IFI, and standardized RMR indices. Most observed indicators also showed acceptable standardized loadings, and the overall construct reliability coefficient reached 0.82. At the same time, the weaker loading of SF4, the low AVE and CR values across the first-order dimensions, and the improper second-order estimate for Strategic Flexibility indicate that the evidence does not yet justify an unrestricted claim of complete construct validity.

Overall, the results offer preliminary support for the general structure of the MCPS assessment while also revealing areas that require further improvement. In particular, the Strategic Flexibility items should be revised, more indicators should be added to each first-order factor, and the hierarchical measurement model should be tested again using a larger and more heterogeneous sample.

DISCUSSION

The psychometric findings provide preliminary support for the use of the instrument in measuring mathematical creative problem-solving (MCPS) as a multidimensional construct. The content-validation stage yielded Aiken's V coefficients between 0.85 and 0.95, reflecting a high level of expert agreement concerning the relevance and representativeness of the assessment items. The second-order CFA also supported the proposed hierarchical structure, with Numerical Fluency, Strategic Flexibility, Structural Innovation, and Analytical Integration functioning as first-order dimensions of the broader MCPS construct. At the model level, the fit statistics were satisfactory, as shown by the non-significant Chi-square value, an RMSEA of 0.041, and strong GFI, AGFI, NNFI, CFI, and IFI values. Those values suggest the reasonable correspondence of the model with the empirical data and offer an acceptable basis for assessing mathematical creative problem-solving. Recent research on instrument validation also found the excellence of construct evidence due to the combination between expert-based content validation and confirmatory factor analysis compared to a single-source evidence interpretation (Boateng et al., 2018; Kyriazos, 2018; Morgado et al., 2017).

Simultaneously, the item-level results found several aspects with the necessities to refine. The score of weaker loading is correlated to strategic flexibility, indicating any other possible indicators to provide benefit. This result also suggests a wiser range of strategic adaptation in mathematical problem-solving. Previous studies also argue the challenges of strategic flexibility operationalization than the procedural fluency due to the required modification, comparison, and evaluation upon multiple solution routes instead of merely carrying out familiar procedures (Meier et al., 2021; Schoevers et al., 2020). Although the overall Construct Reliability yields a satisfactory category (CR = 0.82), the modest AVE values for several first-tier dimensions indicate the urgency of further indicator revisions to improve convergent validity. However, this finding does not merely indicate the dimensions' shortcomings. Instead, this finding reflects the iterative nature of instrument development. Therefore, it requires empirical examination to evaluate the proposed construct and identify items or dimension to further refine before broader implementation (Lambert & Newman, 2023; Lim et al., 2026).

The results indicate that combining a digital platform with a two-tier assessment structure provides a broader basis for evaluating mathematical creative problem-solving than conventional mathematics tests. Traditional assessments tend to prioritize whether students arrive at the correct final answer, leaving much of the reasoning behind that answer unexamined. The instrument developed in this study addresses this limitation by linking selected-response items in Tier 1 with open-ended explanations in Tier 2. Through this arrangement, the assessment captures both the accuracy of students' solutions and the cognitive processes involved in producing them. It therefore offers more detailed evidence of conceptual understanding, strategic

choice, and creative mathematical reasoning, consistent with contemporary perspectives that view mathematical competence as encompassing the formulation, representation, justification, and evaluation of solution strategies rather than merely producing correct answer (Botelho et al., 2023; Copur-Gencturk & Doleck, 2021; Mukuka et al., 2023).

Its digital delivery also enhanced the practical implementation of the assessment. The web-based system administered the items, randomized their sequence, recorded student responses, and scored Tier 1 automatically, while retaining Tier 2 explanations for rubric-based assessment. This configuration not only streamlined administration but also produced a richer record of students' mathematical thinking than would normally be available from paper-based testing. Earlier studies have likewise shown that digital assessment systems can support systematic collection of process evidence, accelerate feedback, and strengthen educational decision-making by documenting different dimensions of student performance (Ifenthaler & Gibson, 2020; Morze et al., 2023; Redecker & Johannessen, 2013; Zainuddin et al., 2026). In this sense, the present model uses digital technology for more than test delivery. It connects automated scoring with structured analysis of mathematical reasoning, thereby supporting process-oriented assessment of higher-order mathematical competencies.

The current study also provides further contribution in terms of incorporating ethnomathematical settings into near-transfer and far-transfer tasks for assessment MCPS. In this research, the researchers did not perceive the cultural elements as background illustrations but as authentic practices of Central Java, embedded directly into the mathematical structure of the problems. Therefore, the students had to interpret realistic situations, identify relevant mathematical relationship, and generate multiple possible solution strategy. The current assessment design accommodates students to relate formal mathematical ideas to culturally meaningful experiences while engaging in creative reasoning. Previous studies also found the superiority of ethnomathematical and culturally situated approached to improve students' conceptual understanding by connecting abstract mathematical ideas and familiar cultural practices. Thus, students would experience relevant mathematical activity while supporting culturally responsive participation and the development of positive mathematical identities (Garcia-Olp et al., 2022; Hunter & Miller, 2022; Payadnya et al., 2024).

The findings put the current instrument within the latest progression of ethnomathematics research. The integrated cultural elements and mathematics curricula provide a meaningful approach to mathematics education. On the other hand, the routine cultural practices could reveal multiple mathematical connections across disciplinary and sociocultural domains (Fouze & Amit, 2018; Rodríguez-Nieto & Alsina, 2022). In Indonesia, ethnomathematics research has been predominantly qualitative, with a strong concentration of studies on geometry and culturally embedded artefacts (Turmuzi et al., 2023; Wibawa et al., 2025). More recently, studies incorporate ethnomathematical connection into didactic sequence to support mathematical problem-solving within culturally meaningful contexts (Rodríguez-Nieto et al., 2026). The current research contributes to the assessment direction by using cultural practices as mathematically meaningful tasks contexts within a digital two-tier instrument and by examining the proposed construct structure using student-response data. Simultaneously, the recorded challenges while integrating ethnomathematics into classroom instruction indicate the importance of providing clear scoring guidelines and adequate teacher preparation to implement such an instrument, particularly for interpreting open-ended reasoning and ensuring that cultural contexts remain mathematically authentic (Sunzuma & Maharaj, 2019).

The use of both near-transfer and far-transfer tasks also expanded the assessment's capacity to capture different forms of mathematical reasoning. Near-transfer items asked students to apply previously learned concepts to culturally familiar situations that retained structural similarities to earlier learning experiences. Far-transfer tasks, by contrast, required students to reconstruct prior knowledge and devise new strategies in less familiar, non-routine contexts. Combining these two forms of transfer offers a broader representation of mathematical creativity because students must do more than reproduce established procedures; they are also expected to adapt, reorganize, and extend their knowledge when solving unfamiliar problems. This interpretation is consistent with research that positions transfer as an important indicator of deep mathematical understanding and creative problem-solving, especially when learners must connect prior knowledge with authentic situations beyond the classroom (Nathan & Alibali, 2021; Orón & Lizasoain, 2023; Sala & Gobet, 2019). Accordingly, the integration of ethnomathematics and transfer-oriented tasks provides a culturally responsive basis for examining how flexibly students apply mathematical knowledge across varied problem contexts.

This study constitutes a further stage of the instrument-development work initiated by Ulya et al. (2024a), rather than a repetition of that study. The earlier research focused on defining the MCPS construct, formulating its four dimensions, and establishing content validity through expert judgment. It also introduced real-world stimuli, divergent and convergent thinking, SOLO taxonomy levels, and near- and far-transfer tasks as the conceptual basis of the instrument. However, its evidence was limited to expert appraisal and did not test the proposed structure using student-response data.

The present study extends that foundation in three ways. First, the instrument was converted into a digital two-tier format that links selected answers with open-ended mathematical reasoning. Second, ethnomathematical contexts were embedded more directly in non-routine tasks so that cultural practices became part of the mathematical demand. Third, the proposed structure was tested with responses from 240 eighth-graders using second-order CFA. This analysis examined whether Numerical Fluency, Strategic Flexibility, Structural Innovation, and Analytical Integration functioned as first-order dimensions of the broader MCPS construct. The findings provide preliminary support for the overall framework while also revealing psychometric limitations that could not be identified through content validation alone. Thus, the main contribution of this study lies in advancing the original prototype into a digitally administered and empirically examined assessment.

Several limitations should be acknowledged. Although the global fit indices were satisfactory, the evidence at the dimension level was weaker. SF4 showed the lowest loading, and Strategic Flexibility produced comparatively low AVE and CR values. Its second-order loading also exceeded 1.00 and was accompanied by a negative residual variance, indicating a Heywood-type estimation problem. These results mean that satisfactory global fit should not be interpreted as full confirmation of the

hierarchical model. The use of only two indicators for each first-order factor may also have limited the stability and breadth of measurement.

The study involved 240 students from two schools within one geographical area, so the results may reflect local curriculum practices, cultural familiarity, school characteristics, and students' experience with digital assessment. Tier 2 responses were also scored by a single trained evaluator; therefore, inter-rater reliability could not be examined. The instrument should consequently be regarded as providing preliminary validation evidence rather than as a finalized measure for unrestricted use.

Even so, the assessment has practical value for mathematics teachers. Its two-tier format helps distinguish correct answers supported by sound reasoning from those produced through guessing or procedural imitation. The open-ended component can also reveal alternative strategies, misconceptions, and difficulties in transferring mathematical knowledge, allowing teachers to provide more targeted feedback. In addition, ethnomathematical tasks can make assessment more culturally meaningful by connecting formal concepts with familiar situations.

Future studies should refine or replace the Strategic Flexibility items and increase the number of indicators for each dimension. The revised model should be tested with larger and more diverse samples through CFA, cross-validation, and measurement-invariance analysis. Further research should also involve multiple raters, report inter-rater reliability, and examine response processes through interviews or think-aloud protocols. Cross-cultural testing would help determine whether the instrument measures MCPS consistently across different ethnomathematical contexts.

CONCLUSION

This study developed an ethnomathematics-based digital two-tier assessment designed to capture both the outcomes and reasoning processes involved in students' mathematical creative problem-solving. The instrument integrates complex multiple-choice items in Tier 1 with open-ended explanations in Tier 2, while embedding culturally meaningful, non-routine tasks that support divergent and convergent thinking as well as near- and far-transfer. In this way, the study addresses its first objective by providing a structured assessment model that links mathematical accuracy, strategic reasoning, and ethnomathematical context within a single digital platform.

The psychometric analysis provided preliminary support for the overall MCPS measurement framework. The model demonstrated satisfactory global fit, and the complete instrument achieved an overall construct reliability coefficient of 0.82. However, the weaker loading of SF4, the comparatively low AVE and CR values at the dimension level, and the Heywood-type estimate associated with Strategic Flexibility indicate that the hierarchical structure should be interpreted cautiously. Accordingly, the instrument is currently better supported as a composite measure of MCPS than as four fully independent subscales.

For practice, the assessment can help mathematics teachers distinguish correct answers supported by coherent reasoning from responses produced through guessing or procedural imitation. Its open-ended component may also reveal alternative strategies, misconceptions, and difficulties in transferring mathematical knowledge, while the ethnomathematical contexts make assessment more meaningful and culturally responsive. Future studies should refine the Strategic Flexibility items, increase the number of indicators within each dimension, involve multiple raters, and test the revised model with larger, more diverse, and cross-cultural samples.

Author contributions: HU: conceptualization, methodology, instrument development, data collection, formal analysis, writing – original draft, funding support; **RR:** conceptualization, instrument development, data collection, providing technical support, supervision, funding support; **ACM:** conceptualization, design, software, providing technical support, funding support; **ABB:** methodology, writing - review & editing, supervision. All authors have sufficiently contributed to the study, and agreed with the results and conclusions.

Funding: The authors would like to sincerely acknowledge the Directorate of Research and Community Service, Directorate General of Research and Development, and the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia for their financial support through the 2025 research grant, under master contract number 127/C3/DT.05.00/PL/2025, derivative contract number 008/LL6/PL/AL.04/2025, and agreement number 275/LPPM.UMK/B.07.06/VI/2025.

Ethical statement: The authors stated that the study was approved by the Research Ethics Committee of Universitas Muria Kudus under Ethical Approval No. 292/LPPM.UMK/B.09.129/VI/2025. Written informed consents were obtained from the participants.

AI statement: The authors stated that generative artificial intelligence (AI) tools were used solely to improve the English language, grammar, and linguistic clarity of the manuscript. All scientific content, research analyses, interpretations, original ideas, and conclusions were independently developed, critically reviewed, and verified by the authors, who retain full responsibility for the accuracy, integrity, and originality of the manuscript.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement, 40*(4), 955–959. https://doi.org/10.1177/001316448004000419

Antika, H. N., Anjasari, T., & Kohar, A. W. (2023). Students' creativity in solving ill structured problem context calorie intake. *Journal of Mathematical Pedagogy, 4*(1), 42–56. https://doi.org/10.26740/jomp.v4n1.p42-56

Asamoah, D. B. (2019). Traditional assessment procedures, and performance and portfolio assessment procedures: An in-depth comparison. *International Journal of Educational Research and Studies, 1*(2), 28–30.

Azmi, N., Arif, S., Sofyan, H., & Oktavia, R. (2025). Bridging geometry and cultures for junior high school level: Rumoh Aceh design from a computational thinking perspective. *Journal on Mathematics Education, 16*(2), 383–406. https://doi.org/10.22342/jme.v16i2.pp383-406

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Bicer, A. (2021). Multiple representations and mathematical creativity. *Thinking Skills and Creativity, 42*(December). https://doi.org/10.1016/j.tsc.2021.100960

Bicer, A., Aldemir, T., Krall, G., Quiroz, F., Chamberlin, S., Nelson, J. L., Lee, Y., & Kwon, H. (2024). Exploring creativity in mathematics assessment: An analysis of standardized tests. *Thinking Skills and Creativity, 54*(December). https://doi.org/10.1016/j.tsc.2024.101652

Boaler, J., Brown, K., LaMar, T., Leshin, M., & Selbach-Allen, M. (2022). Infusing mindset through mathematical problem solving and collaboration: Studying the impact of a short college intervention. *Education Sciences, 12*(10), Article 694. https://doi.org/10.3390/educsci12100694

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Borg, W. R., & Gall, M. D. (1989). *Educational research: An introduction* (5th ed.). Longman.

Botelho, A., Baral, S., Erickson, J. A., Benachamardi, P., & Heffernan, N. T. (2023). Leveraging natural language processing to support automated assessment and feedback for student open responses in mathematics. *Journal of Computer Assisted Learning, 39*(3), 823–840. https://doi.org/10.1111/jcal.12793

Copur-Gencturk, Y., & Doleck, T. (2021). Strategic competence for multistep fraction word problems: an overlooked aspect of mathematical knowledge for teaching. *Educational Studies in Mathematics, 107*, 49–70. https://doi.org/10.1007/s10649-021-10028-1

D'Ambrosio, U. (1985). Ethnomathematics and its place in the history and pedagogy of mathematics. *For the Learning of Mathematics, 5*(1), 44–48.

Faulkner, F., Breen, C., Prendergast, M., Carr, M., Faulkner, F., Breen, C., Prendergast, M., & Carr, M. (2023). Profiling mathematical procedural and problem- solving skills of undergraduate students following a new mathematics curriculum. *International Journal of Mathematical Education In, 54*(2), 220–249. https://doi.org/10.1080/0020739X.2021.1953625

Fouze, A. Q., & Amit, M. (2018). On the importance of an ethnomathematical curriculum in mathematics education. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(2), 561–567. https://doi.org/10.12973/ejmste/76956

Garcia-Olp, M., Nelson, C., & Saiz, L. (2022). Decolonizing mathematics curriculum and pedagogy: Indigenous knowledge has always been mathematics education. *Educational Studies, 58*(1), 1–16. https://doi.org/10.1080/00131946.2021.2010079

Hamid, N. H. A., & Kamarudin, N. (2020). Cultivating mathematical creativity among students: Learning for 21st century education. *International Research Journal of Education and Sciences (IRJES), 4*(2), 5–9.

Hirsh, Å. (2020). When assessment is a constant companion: students' experiences of instruction in an era of intensified assessment focus. *Nordic Journal of Studies in Educational Policy, 6*(2), 89–102. https://doi.org/10.1080/20020317.2020.1756192

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education*. Center for Curriculum Redesign.

Hunter, J., & Miller, J. (2022). The use of cultural contexts for patterning tasks: supporting young diverse students to identify structures and generalise. *ZDM – Mathematics Education, 54*(6), 1349–1362. https://doi.org/10.1007/s11858-022-01386-y

Hurrell, D. P. (2021). Conceptual knowledge OR Procedural knowledge OR Conceptual knowledge AND Procedural knowledge: Why the conjunction is important for teachers. *Australian Journal of Teacher Education, 46*(2), 57–71. https://doi.org/10.14221/ajte.2021v46n2.4

Ifenthaler, D., & Gibson, D. (2020). *Adoption of data analytics in higher education learning and teaching*. Springer. https://doi.org/10.1007/978-3-030-47392-1

Kattou, M., Kontoyianni, K., Pitta-Pantazi, D., & Christou, C. (2012). Connecting mathematical creativity to mathematical ability. *ZDM - International Journal on Mathematics Education, 45*(2), 167–181. https://doi.org/10.1007/s11858-012-0467-1

Kaya, D., & Yavuz, S. (2025). Evaluation of creative activities prepared with an ethnomathematics approach from the perspective of realistic mathematics education principles. *Irish Educational Studies*, 1–23. https://doi.org/10.1080/03323315.2025.2546964

Keleş, T. (2022). Investigation of high school students' creative problem-solving attributes. *Journal of Pedagogical Research, 6*(4), 66–83. https://doi.org/10.33902/JPR.202215433

Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA , CFA) and SEM in general. *Psychology, 9*, 2207–2230. https://doi.org/10.4236/psych.2018.98126

Lambert, L. S., & Newman, D. A. (2023). Construct development and validation in three practical steps : Recommendations for reviewers, editors, and authors. *Feature Topic: Reviewer Resources, 26*(4), 574–607. https://doi.org/10.1177/10944281221115374

Lee, K. S., Hwang, D., & Seo, J. J. (2003). A development of the test for mathematical creative problem solving ability. *Journal of the Korea Society of Mathematical Education, Series D: Research in Mathematical Education, 7*(3), 163–189.

Leikin, R., & Pitta-Pantazi, D. (2013). Creativity and mathematics education: The state of the art. *ZDM - International Journal on Mathematics Education, 45*(2), 159–166. https://doi.org/10.1007/s11858-012-0459-1

Lim, W. M., Khatri, P., Shiva, A., & Dabas, V. (2026). Guidelines for scale development and validation. *Journal of Consumer Behaviour, 25*, 953–973. https://doi.org/10.1002/cb.70096

Mei, A., Marsigit, Purwastuti, L. A., Hidayat, R., & Ayub, A. F. M. (2025). Problem-solving learning model based on ethnomathematics to improve student's creative thinking in elementary school. *Malaysian Journal of Mathematical Sciences, 19*(1), 289–309. https://doi.org/10.47836/mjms.19.1.15

Meier, M. A., Burgstaller, J. A., Benedek, M., Vogel, S. E., & Grabner, R. H. (2021). Mathematical creativity in adults: Its measurement and its relation to intelligence, mathematical competence and general creativity. *Journal of Intelligence, 9*(1), Article 10. https://doi.org/10.3390/jintelligence9010010

Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica, 30*, Article 3. https://doi.org/10.1186/s41155-016-0057-1

Morze, N., Smyrnova-Trybulska, E., Drlík, M., & Buinytska, O. (2023). Development of advanced digital ecosystems at universities: A study comparing experiences from Ukraine, Poland and Slovakia. *European Journal of Education, 58*, 647–664. https://doi.org/10.1111/ejed.12588

Mukuka, A., Balimuttajjo, S., & Mutarutinya, V. (2023). Teacher efforts towards the development of students' mathematical reasoning skills. *Heliyon, 9*(4), Article e14789. https://doi.org/10.1016/j.heliyon.2023.e14789

Nathan, M. J., & Alibali, M. W. (2021). An embodied theory of transfer of mathematical learning. In C. Hohensee & J. Lobato (Eds.), *Transfer of learning: Progressive perspectives for mathematics education and related fields* (pp. 27–58). Springer.

Orey, D., & Rosa, M. (2007). Cultural assertions and challenges towards pedagogical action of an ethnomathematics program. *For the Learning of Mathematics, 27*(1), 10–16.

Orón, J. V., & Lizasoain, I. (2023). Achieving transfer from mathematics learning. *Education Sciences, 13*(2), Article 161. https://doi.org/10.3390/educsci13020161

Payadnya, I. P. A. A., Wulandari, I. G. A. P. A., Puspadewi, K. R., & Saelee, S. (2024). The significance of ethnomathematics learning: a cross-cultural perspectives between Indonesian and Thailand educators. *Journal for Multicultural Education, 18*(4), 508–522. https://doi.org/10.1108/JME-05-2024-0049

Redecker, C., & Johannessen, Ø. (2013). Changing assessment-Towards a new assessment paradigm using ICT. *European Journal of Education, 48*(1), 79–96. https://doi.org/https://doi.org/10.1111/ejed.12018

Rodríguez-Nieto, C. A., & Alsina, Á. (2022). Networking between ethnomathematics, STEAM education, and the globalized approach to analyze mathematical connections in daily practices. *EURASIA Journal of Mathematics, Science and Technology Education, 18*(3), Article em2085. https://doi.org/10.29333/EJMSTE/11710

Rodríguez-Nieto, C. A., Labio-Petefi, K. T., Palacios-Balanta, Y. K., García-Moreno, A., Inostroza, A. C., Cantillo-Rudas, B. M., & Sudirman, S. (2026). Fostering arithmetic problem-solving in primary education through didactic sequences with ethnomathematical connections using chontaduro commercialization. *EURASIA Journal of Mathematics, Science and Technology Education, 22*(9), Article em2899. https://doi.org/10.29333/ejmste/19123

Rosa, M., D'Ambrosio, U., Orey, D. C., Shirley, L., Alangui, W. V., Palhares, P., & Gavarrete, M. E. (2016). *Current and future perspectives of ethnomathematics as a program*. Springer.

Sala, G., & Gobet, F. (2019). Cognitive training does not enhance general cognition. *Trends in Cognitive Sciences, 23*(1), 9–20.

Schoevers, E. M., Kroesbergen, E. H., & Kattou, M. (2020). Mathematical creativity: A combination of domain- general creative and domain-specific mathematical skills. *Journal of Creative Behavior, 54*(2), 242–252. https://doi.org/10.1002/jocb.361

Silver, E. A. (1997). Fostering creativity through instruction rich in mathematical problem solving and problem posing. *Zentralblatt Für Didaktik Der Mathematik, 29*(3), 75–80. https://doi.org/10.1007/s11858-997-0003-x

Sipahi, Y., & Bahar, A. K. (2025). Mathematical creativity: A systematic review of definitions, frameworks, and assessment practices. *Education Sciences, 15*(10), Article 1348. https://doi.org/10.3390/educsci15101348

Suherman, S., & Vidákovich, T. (2022). Assessment of mathematical creative thinking: A systematic review. *Thinking Skills and Creativity, 44*, Article 101019. https://doi.org/10.1016/j.tsc.2022.101019

Sunzuma, G., & Maharaj, A. (2019). Teacher-related challenges affecting the integration of ethnomathematics approaches into the teaching of geometry. *EURASIA Journal of Mathematics, Science and Technology Education, 15*(9), Article em1744. https://doi.org/10.29333/ejmste/108457

Susanti, D., & Yuhendri, L. V. (2024). The effectiveness of e-assessment in improving the quality of learning and the quality of assessment on financial accounting learning: a literature review. *Journal of Education, Teaching, and Learning, 9*(1), 22–28. https://doi.org/10.26737/jetl.v9i1.5911

Sutarto, Muzaki, A., Hastuti, I. D., Fujiaturrahman, S., & Untu, Z. (2022). Development of an ethnomathematics-based e-module to improve students' metacognitive ability in 3D geometry topic. *International Journal of Interactive Mobile Technologies , 16*(3), 32–46. https://doi.org/10.3991/IJIM.V16I03.24949

Turmuzi, M., Suharta, I. G. P., & Suparta, I. N. (2023). Ethnomathematical research in mathematics education journals in Indonesia: A case study of data design and analysis. *EURASIA Journal of Mathematics, Science and Technology Education, 19*(1), Article em2220. https://doi.org/10.29333/ejmste/12836

Ulya, H., & Rahayu, R. (2021). Students' Mathematical Representation Ability in Kudus Local Wisdom-Based Open-Ended Learning. *Journal of Physics: Conference Series, 1823*, Article 012102. https://doi.org/10.1088/1742-6596/1823/1/012102

Ulya, H., Sugiman, & Rosnawati, R. (2024a). Design and content validity of mathematics creative problem-solving ability instrument for junior high school students. *EURASIA Journal of Mathematics, Science and Technology Education, 20*(6), Article em2462. https://doi.org/10.29333/ejmste/14661

Ulya, H., Sugiman, Rosnawati, R., & Retnawati, H. (2024b). Technology-based learning interventions on mathematical problem-solving: a meta-analysis of research in Indonesia. *International Journal of Evaluation and Research in Education, 13*(1), 292–301. https://doi.org/10.11591/ijere.v13i1.26380

Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & van Tartwijk, J. (2020). Creative problem solving in primary education: Exploring the role of fact finding, problem finding, and solution finding across tasks. *Thinking Skills and Creativity, 37*(April), Article 100665. https://doi.org/10.1016/j.tsc.2020.100665

Wibawa, F. S., Hariri, D. D., Mahmudah, H., & Kania, N. (2025). Bridging mathematics and culture: A systematic review of indigenous mathematical concepts in Indonesian traditions. *Contemporary Mathematics and Science Education, 6*(2), Article ep25013. https://doi.org/10.30935/conmaths/17076

Zainuddin, A., Wasis, Ekohariadi, Anistyarsari, Y., Rijanto, T., & Habibi. (2026). A systematic review of digital assessment trends in education from 2015-2025. *Multidisciplinary Reviews, 9*, Article e2026433. https://doi.org/10.31893/multirev.2026433