

## VALIDITY AND RELIABILITY OF CREATIVE THINKING SKILLS ASSESSMENT INSTRUMENTS IN MATHEMATICS SUBJECTS FOR ELEMENTARY SCHOOL

Mia Hafizah Tumangger<sup>1</sup>, Ryzka Cahyaningrum<sup>2</sup>, Anik Maghfiroh<sup>3</sup>, Sahrul<sup>4</sup>

<sup>1-3</sup> Universitas Negeri Semarang, Indonesia

<sup>4</sup> Universitas Ivet Semarang, Indonesia

[miahafizahtumangger30@mail.unnes.ac.id](mailto:miahafizahtumangger30@mail.unnes.ac.id)

**ABSTRACT** Creative thinking is an essential skill that needs to be developed from an early age. The absence of valid and reliable instruments poses major challenges in identifying and enhancing students' creative thinking skills in elementary school mathematics learning. This study aimed to evaluate the validity and reliability of a creative thinking assessment instrument for elementary school mathematics subjects. A quantitative approach was employed with five expert validators and 150 research participants selected through total sampling. The research instruments included expert validation sheets and a creative thinking assessment instrument for elementary mathematics. Data were analyzed using several methods: Aiken's V for content validity, Rasch model analysis for dimensionality, Confirmatory Factor Analysis (CFA) for construct validity, Ebel's formula for inter-rater reliability, and Cronbach's Alpha for internal consistency reliability. The results showed that the creative thinking assessment instrument was valid and reliable, indicating that it can be used as an accurate measurement tool to identify students' creative thinking skills. The findings of this study serve as a valuable reference for educators in evaluating creative thinking and provide a foundation for developing more innovative learning strategies in mathematics education.

**Keywords:** creative thinking, validity, reliability, assessment, elementary mathematics

**ABSTRAK** Berpikir kreatif merupakan keterampilan esensial yang perlu dikembangkan sejak dini. Ketiadaan instrumen yang valid dan reliabel menjadi tantangan utama dalam mengidentifikasi serta mengembangkan kemampuan berpikir kreatif siswa pada pembelajaran matematika di sekolah dasar. Penelitian ini bertujuan untuk menilai validitas dan reliabilitas instrumen asesmen kemampuan berpikir kreatif pada mata pelajaran matematika sekolah dasar. Penelitian ini menggunakan pendekatan kuantitatif dengan melibatkan lima validator dan 150 sampel penelitian yang ditentukan melalui teknik *total sampling*. Instrumen penelitian terdiri atas lembar validasi ahli dan instrumen asesmen berpikir kreatif matematika sekolah dasar. Analisis data dilakukan dengan beberapa metode, yaitu Aiken's V untuk validitas isi, *Rasch model* untuk uji dimensi, *Confirmatory Factor Analysis*

(CFA) untuk validitas konstruk, rumus Ebel untuk reliabilitas antarpemilai, serta *Cronbach's Alpha* untuk reliabilitas internal. Hasil penelitian menunjukkan bahwa instrumen asesmen kemampuan berpikir kreatif tergolong valid dan reliabel, sehingga dapat digunakan sebagai alat ukur yang akurat untuk mengidentifikasi kemampuan berpikir kreatif siswa. Temuan penelitian ini dapat dijadikan acuan bagi pendidik dalam mengevaluasi kemampuan berpikir kreatif sekaligus sebagai dasar dalam pengembangan strategi pembelajaran yang lebih inovatif.

**Kata-kata kunci:** berpikir kreatif, validitas, reliabilitas, asesmen, matematika sekolah dasar

## INTRODUCTION

Education must cultivate individuals prepared to confront the challenges of the 21st century with the relevant skills and competencies (Slamet & Wardani, 2020). These competencies are instrumental in empowering students to confront challenges in the educational environment and cultivate autonomy and self-efficacy (Barta et al., 2022). Furthermore, the ability to solve problems effectively, make sound decisions, and recognize personal competence in thinking is important (Kavenuke et al., 2020). The 21st-century skills framework is categorized into four pillars: higher-order thinking skills, communication skills, collaboration skills, and creative thinking skills (Mu'minah, 2021). Creativity has been identified as a significant component in the early years of schooling, with subsequent increases in school rankings that underscore the importance of creative thinking skills in education (Umar & Ahmad, 2019). Creativity is also part of the learning taxonomy widely adopted in educational practice.

From higher to basic levels, creative education plays a pivotal role in fostering student motivation to learn (Nakano & Wechsler, 2018). Basic education plays a fundamental role in developing creative thinking skills in children (Susanti et al., 2023). Creative thinking has been proven to encourage children's problem-solving abilities, which are measured based on their ability levels (Tumangger et al., 2022). The cultivation of creative thinking skills has been shown to facilitate children's ability to think fluently and flexibly, understand problems from various perspectives, and generate numerous ideas. Therefore, cultivating creative thinking skills must be initiated and systematically integrated into the elementary education curriculum.

Creative thinking skills can be introduced in elementary education through mathematics lessons using two-dimensional shapes (T et al., 2020). Mathematics learning requires a deep understanding to answer each question, so creative thinking skills are necessary for comprehending mathematics lessons (Y. Putra, 2020). However, only 12.88% of students demonstrated these skills, indicating a significant deficiency (Meika & Sujana, 2017). Lack of attention during math learning can limit creativity (H. D. Putra et al., 2018). Improving these skills also depends heavily on students (Amhar et al., 2022). Teachers and students must collaborate to foster and enhance students' creative thinking skills in math.

Furthermore, it is imperative to cultivate mathematical creative thinking skills through deliberate training to ensure that students become proficient in naturally applying these skills. One method by which students can hone their mathematical aptitude is by solving mathematical problems. Math exercises can improve subsequent mathematical performance (Avvisati & Borgonovi, 2020). Educators can assess students' creative thinking abilities through their responses (Hidayah et al., 2021). Furthermore, educators may evaluate students' abilities to enhance the efficacy of mathematics education. However, the quality of the questions must be considered to ensure that student assessment and practice questions are carried out correctly in accordance with the principles of assessment instruments and creative thinking skills. Some teachers have not constructed assessment instruments that effectively measure creative thinking skills (Haryanti & Saputra, 2019). In fact, using a creative approach to solve math questions is more effective in improving mathematical understanding than simply following routine procedures that students often learn by rote (Jonsson et al., 2020). Therefore, teachers and students need the right assessment instrument to measure creative thinking skills in mathematics.

The quality of assessment instruments is evaluated based on their validity and reliability in measuring mathematical creative thinking skills. Preliminary studies for this study indicated that teachers do not yet have a validated and reliable assessment instrument for mathematical creative thinking skills. The assessment of creative thinking encompasses four domains: fluency, flexibility, originality, and elaboration. Teachers have indicated that the assessment instruments currently in use lack valid and reliable questions that can be used to evaluate students' creative thinking skills in mathematics. It constitutes a significant gap, given that the absence of an instrument with demonstrated validity and reliability can compromise the assessment results, rendering them biased, inconsistent, and inadequate as a foundation for decision-making regarding improvements to learning. Consequently, there is a necessity for the development, testing, and refinement of an assessment instrument capable of accurately and consistently measuring students' creative thinking skills. This instrument would serve to support the enhancement of the quality of mathematics education in elementary schools.

This study examines the validity and reliability of creative thinking skills assessment instruments in elementary mathematics lessons focused on two-dimensional shapes material. The existence of valid and reliable mathematical creative thinking skills assessment instruments for elementary school students is expected to equip and accustom students to think creatively from an early age.

## **METHODS**

---

This study is part of a quantitative research and development project focused on the testing phase, namely the validity and reliability of instruments. The instrument was

developed using the Borg & Gall model to assess elementary school students' creative thinking abilities in mathematics, specifically in the two-dimensional shapes material. The assessment of mathematical creative thinking abilities is predicated on four indicators: fluency, flexibility, originality, and elaboration. Each indicator comprises two essay questions, yielding a total of eight essay questions (FLU1, FLU2, FLE3, FLE4, KSL5, KSL6, KRN7, KRN8) incorporated into the elementary school mathematical creative thinking assessment instrument.

This study involved five validators: two evaluation experts, one mathematics expert, and two practitioners. In addition, this study involved 150 fourth-grade elementary school students from four elementary schools in Besitang Subdistrict, Langkat Regency, which were determined based on total sampling. The instruments used were validator sheets and instruments for assessing elementary school mathematical creative thinking skills.

The validity demonstrated in this study encompasses content validity and construct validity. The instrument's content validity was assessed through Aiken's V, a statistical method based on expert assessment of the relationship between the items and indicators. The Aiken's V value obtained was then interpreted, where a value  $\geq 0.80$  indicates a high level of content validity (Aiken, 1980; Retnawati, 2017). The construct validity of the model was confirmed through dimensionality analysis employing the Rasch model and Confirmatory Factor Analysis (CFA). The dimensionality of the Rasch model was examined to assess the extent to which the instrument items measured the same construct, which was identified as creative thinking skills. The values considered in this analysis are raw variance explained by measures, which ideally exceeds 20%, and unexplained variance in contrast, which does not exceed 15% (Sumintono & Widhiarso, 2014). The interpretation of the results refers to the categories for raw variance explained by measures. Specifically, a value of  $\geq 20\%$  is considered to meet the minimum dimensionality requirement, a value of  $\geq 40\%$  indicates good quality, and a value of  $\geq 60\%$  indicates excellent quality (Hakiki et al., 2018). This study employed Rasch model dimensionality analysis using Winsteps software.

A CFA analysis was performed to verify the alignment of the theoretical construct with the empirical data, specifically the indicators of creative thinking skills and the items of the assessment instrument. The analysis referred to the model fit criteria, which include standardized loading factors of at least 0.50, a Comparative Fit Index (CFI) of at least 0.90, a Goodness of Fit Index (GFI) of at least 0.90, a Normed Fit Index (NFI) of at least 0.90, a Root Mean Square Error of Approximation (RMSEA) of less than 0.08, and a Chi-Square/df ratio of less than 3.0 (Cai et al., 2021; Hair Jr et al., 2019; Hu & Bentler, 1999; Kline, 2016; Narimawati et al., 2020). This study employed LISREL software to conduct the analysis.

The reliability of the study was estimated using two approaches: inter-rater reliability and internal reliability. The inter-rater reliability of the assessments was estimated

using Ebel's formula to ascertain the consistency of the inter-rater assessments. Ebel's reliability values are interpreted as follows: categories  $\geq 0.75$  indicate high levels, categories between 0.60 and 0.74 indicate moderate levels, and categories  $< 0.60$  indicate low levels. Concurrently, the internal reliability of the instrument was gauged employing Cronbach's Alpha. An alpha value of  $\geq 0.70$  indicates acceptable reliability, while a value of  $\geq 0.80$  indicates a high level of reliability (Sugiyono, 2019).

## FINDING AND DISCUSSION

### Content Validity

This study initiated its data analysis by assessing content validity, as determined by the results of the validator sheet. The validator sheet contains eight statements, each assigned a scale or score ranging from 1 to 4. A score of 1 indicates that the assessment instrument item is not aligned with the indicator. A score of 2 indicates that the item is somewhat aligned with the indicator. A score of 3 indicates that the item is aligned with the indicator. A score of 4 indicates that the item is highly aligned with the indicator (Azwar, 2021). A total of five validators contributed to the completion of the validator sheet. The results of the validator sheet were subsequently analyzed using Aiken's V. The results are shown in Table 1.

Table 1. Content Validity

Item	Aiken's V	Category
1	0.93	Valid
2	0.85	Valid
3	0.85	Valid
4	0.93	Valid
5	0.93	Valid
6	0.93	Valid
7	0.93	Valid
8	0.93	Valid

As demonstrated in Table 1, all items present Aiken's V values greater than 0.80, thereby substantiating the validity of the entire set of items. However, the validator provided editorial notes regarding item four's lack of communicative instructions. Conversely, the assessment instrument aligns with elementary school mathematics's conceptual framework and material and creative thinking skills indicators. Consequently, the researcher modified the wording of the item without altering its underlying content and proceeded with the validity analysis of the remaining items.

### Inter-Rater Reliability

The subsequent data analysis will address the issue of inter-rater reliability. The inter-rater reliability procedure is conducted to estimate the level of consistency among validators in providing assessments on the assessment instrument. The results of the inter-rater reliability analysis support the assessment instrument's validity. The reliability of the system is analyzed using the Ebel formula. As illustrated in Table 2, the results of the inter-rater reliability analysis are presented.

**Table 2.** Inter-Rater Reliability

Parameter	Value
Mean Square Residual	0.302
Mean Square Between Raters	2.088
Reliability Coefficient (Rxx)	0.855
Interpretation	High

The results of the calculations, performed using the Ebel formula, yielded an inter-rater reliability value of 0.855, which is classified as high. Accordingly, the validators showed consistent agreement, thereby confirming the reliability of the assessment instrument in measuring elementary school mathematical creative thinking skills.

### Construct Validity

Construct validity is the subsequent feasibility assessment for research instruments determined to be valid and reliable based on validator data. The data employed in the construct validity assessment are derived from assessment instrument data collected from a sample of fourth-grade elementary school students. Construct validity aims to assess the dimensionality and appropriateness of the assessment instrument construct.

### Dimensionality

The dimensionality of construct validity is critical in ensuring that the instrument items measure one primary construct or indicate additional dimensions. The dimensionality test was conducted using dimensionality analysis in the Rasch model. The results of the dimensionality test are presented in Table 3.

**Table 3.** Dimensionality Test

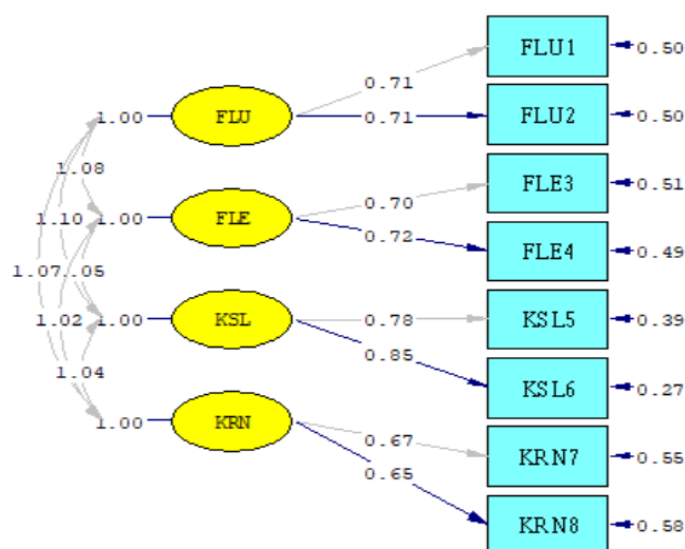
Parameter	Eigenvalue/ Observed
Total raw variance in observations	28.95 (100%)
Raw variance explained by measures	20.95 (72.4%)
Raw explained variance (total)	8.00 (27.6%)
Unexplained variance in 1st contrast	3.61 (12.5%)

Note. Contrast is part of the residual variance, not an additional source of variance.

The dimensionality test results indicated several significant parameters. The total observed value was 100%. The total raw variance in observations was then divided into two categories: raw variance explained by measures, which accounts for more than 20% of the total, and unexplained variance in the first contrast, which makes up less than 15% of the total. This process served to demonstrate the unidimensional nature of a construct. Based on Table 3, the raw variance explained by the measures had an observed value of 72.4%, greater than 20%, with an eigenvalue of 20.95. Then, the observed value of the unexplained variance in the first contrast parameter was 12.5%, less than 15%. These results indicated that the dimensionality test was fulfilled. Therefore, the assessment instrument was determined to measure the primary, or unidimensional, construct of creative thinking skills.

### Confirmatory Factor Analysis (CFA)

Furthermore, this study employed construct validity with CFA factor analysis. The results of Confirmatory Factor Analysis (CFA) for the measurement model are presented in Figure 1.



Chi-Square=15.91, df=16, P-value=0.45931, RMSEA=0.000

**Figure 1.** Standardized Estimates of Confirmatory Factor Analysis (CFA)

A confirmatory factor analysis (CFA) was conducted to validate and select the suitability of items for measuring four subconstructs or indicators of creative thinking skills. The suitability of items was determined by observing the standardized loading factor arrow. If the standardized loading factor for an item was not met, the item was to be discarded or revised. As illustrated in Figure 1, item FLU1 demonstrated a standardized loading factor value of 0.71. Item FLU2 obtained a standardized loading factor value of 0.71, indicating its suitability for measuring creative thinking skills in the fluency subconstruct. Additionally, items FLE3 and FLE4

yielded values of 0.70 and 0.72, respectively, suggesting that items three and four are appropriate for assessing creative thinking skills in the flexibility subconstruct.

Further analysis revealed that items KSL5 and KSL6 obtained standardized loading factors of 0.70 and 0.85, respectively, suggesting their appropriateness for assessing creative thinking skills in the originality subconstruct. Finally, items KRN7 and KRN8 obtained standardized loading factor values of 0.67 and 0.65, respectively, indicating their suitability for measuring creative thinking skills in the elaboration subconstruct. Therefore, all items meet the standardized loading factor criteria, indicating that the instrument items serve as a metric for assessing creative thinking skills.

Following the assurance that all elements are compatible with the assessment instrument, the model fit is analyzed using the goodness-of-fit index criteria. The purpose of this index is to evaluate the degree to which the model in Figure 1 aligns with the data. If the established criteria are met, the model constructed in the path diagram can provide a satisfactory representation of the empirical data. The goodness-of-fit index results are displayed in Table 4.

**Table 4.** Goodness of Fit Index

Index	Acceptance Range	Value	Status
Chi-Square	p-value > 0.05	0.46	Fit
RMSEA	< 0.08	0.00	Fit
GFI	> 0.90	0.93	Fit
CFI	> 0.90	0.99	Fit
NFI	> 0.90	0.93	Fit
Chi-Square/df ( $x^2/df$ )	< 3.0	0.99	Fit

As illustrated in Table 4, there are several indices with acceptance ranges. The goodness-of-fit index values were obtained through model output analysis using the LISREL software. The results of the model fit analysis indicated that each index meets the acceptance range criteria, with a p-value = 0.46, RMSEA = 0.00, GFI = 0.93, CFI = 0.99, NFI = 0.93, and  $x^2/df$  = 0.99. Accordingly, the assessment instrument is deemed valid in its representation of the subconstruct of creative thinking skills, thereby rendering it suitable for measuring the creative thinking skills of elementary school students in mathematics.

**Internal Consistency Reliability**

The validity of the assessment instrument was subsequently estimated based on its internal reliability. The degree to which an instrument can provide consistent results when measuring the same construct is indicated by internal reliability. This study

employed Cronbach's Alpha to assess the reliability of the assessment instrument. The reliability estimation results are presented in Table 5.

**Table 5.** Internal Reliability

Cronbach's Alpha ( $\alpha$ )	N of items	Status
0.85	8	Reliable

A Cronbach's Alpha coefficient with a value of at least 0.70 is generally considered to indicate good reliability. The calculation results indicate that the Cronbach's Alpha value is 0.85. This finding indicates that the instrument presents high internal consistency, ensuring its reliability in measuring creative thinking skills.

The main findings of this study indicate that the instrument used to assess mathematical creative thinking skills with two-dimensional shapes meets the requirements for validity and reliability. According to several studies, assessment instruments should be tested for validity and reliability so that measurements can be carried out optimally and be relevant to the concept (Ahmed & Ishtiaq, 2021; Geschwind et al., 2024; Sullivan, 2011). Febryana et al. (2024) tested the instrument's validity and reliability using the Rasch model and identified responses that were not appropriate for the students. It suggests that validity and reliability are part of evaluating instrument performance.

The results of this study indicated that all items align with the concept of elementary school mathematical creative thinking skills, as evidenced by the content validity results. Furthermore, the results revealed consistent inter-rater reliability. Additionally, the construct validity results provided evidence that supports the feasibility of the mathematical creative thinking skills assessment instrument. The dimensionality test indicated that the instrument is unidimensional, with all items contributing to a single general construct: creative thinking skills. These findings suggest that, in practice, all variables align with the same underlying concept. However, within the theoretical framework, the ability to think creatively comprises four sub-constructs: fluency, flexibility, originality, and elaboration. The CFA results supported this multidimensional model, with model fit indices that meet the established criteria. These results confirmed that the assessment instrument can be regarded as a unifying construct comprising a model of four representative subconstructs, collectively articulating an overall assessment of creative thinking skills. In addition, the internal reliability results indicated that the instrument items consistently measure creative thinking skills.

This assessment instrument is designed for elementary school teachers to identify students' creative thinking skills with two-dimensional shape material in mathematics learning. Moreover, the instrument's functionality extends beyond mere measurement, serving as a medium for creative thinking. Earlier research has indicated that an assessment can function in a dual capacity, serving as both a

measuring instrument and a pedagogical device (Bara & Bara, 2023; Korobko, 2019; Morell & Wilson, 2016). Implementing a validated and reliable instrument is imperative to ensure that the practice process carried out by students through the instrument's questions is genuinely focused on the concept of creative thinking. It is no longer necessary for educators to decide between instructing students on the subject matter of two-dimensional shapes and cultivating creative thinking abilities. These two pursuits can be seamlessly integrated into a unified assessment instrument. Therefore, this instrument can potentially enhance the quality of mathematics learning, focusing more on higher-order thinking skills.

While the study yielded positive results, it is important to note that several other issues were not discussed further and thus became limitations of the study. The evaluation of assessment instruments was initially constrained to two-dimensional shapes and a limited set of samples. To expand the generalizability of the findings, further research could entail adapting and testing related instruments on various mathematical materials and across diverse levels of education, accompanied by more comprehensive analyses. Classroom action research can also be conducted to analyze the instrument's effectiveness when applied continuously by teachers in the learning cycle and to see its impact on improving students' creative thinking skills longitudinally.

## **CONCLUSIONS AND RECOMMENDATIONS**

This study examines the validity and reliability of assessment instruments teachers will use to assess students' creative thinking skills in mathematics. The validity and reliability examined encompass content validity, construct validity, inter-rater reliability, and internal consistency reliability. The content validity results indicate that the assessment instrument items are in accordance with the indicators and concepts developed (0.80 – 0.95), followed by consistent inter-rater reliability (0.855). The dimensionality assessment employing the Rasch model reveals that the instrument manifests unidimensionality, exhibiting an explanatory power of 72.4%, which surpasses the stipulated minimum threshold. Construct validity, as determined by CFA analysis, confirmed the instrument's sub-constructs, exhibiting a satisfactory model suitability index (GFI = 0.93; CFI = 0.99; and NFI = 0.93). Additionally, internal reliability showed excellent instrument consistency ( $\alpha = 0.85$ ). Therefore, the instrument is declared valid and reliable for assessing students' creative thinking abilities in mathematics.

The present study is exclusively concerned with the validity and reliability of the assessment instrument. It is imperative to acknowledge the potential of supplementary studies to thoroughly assess the feasibility of the aforementioned assessment instrument through rigorous item analysis. Furthermore, subsequent research endeavors may investigate the efficacy of the creative thinking assessment instrument in influencing learning outcomes or other variables.

**REFERENCES**

- Ahmed, I., & Ishtiaq, S. (2021). Reliability and validity: Importance in medical research. *JPMA: The Journal of the Pakistan Medical Association*, 71(10), 2401–2406. <https://doi.org/10.47391/JPMA.06-861>
- 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>
- Amhar, A., Sabrina, R., Sulasmi, E., & Saragih, M. (2022). Student critical thinking skills and student writing ability: The role of teachers' intellectual skills and student learning. *International Journal of Evaluation and Research in Education*, 17(7), 2493–2510.
- Avvisati, F., & Borgonovi, F. (2020). Learning mathematics problem solving through test practice: A randomized field experiment on a global scale. *Educational Psychology Review*, 32(3), 791–814. <https://doi.org/10.1007/s10648-020-09520-6>
- Azwar, S. (2021). *Reliabilitas dan validitas* (4th ed.). Pustaka Pelajar.
- Bara, E., & Bara, G. (2023). The connection between assessment and learning outcomes. *Balkan Journal of Interdisciplinary Research*, 9(2), 51–58. <https://doi.org/10.2478/bjir-2023-0006>
- Barta, A., Fodor, L. A., Tamas, B., & Szamoskozi, I. (2022). The development of students' critical thinking abilities and dispositions through the concept mapping learning method. *Educational Research Review*, 37, 100481. <https://doi.org/10.1016/j.edurev.2022.100481>
- Cai, L., Chung, S. W., & Lee, T. (2021). Incremental model fit assessment in the case of categorical data: Tucker–Lewis Index for item response theory modeling. *Prevention Science*, 22(2), 1–9. <https://doi.org/10.1007/s11121-021-01253-4>
- Febryana, U. I., Rahmania, F., Faradillah, A., & Jusra, H. (2024). Evaluation of validity and reliability using Rasch model and SPSS on students' mathematical creative thinking skills. *Pattimura Proceeding: Conference of Science and Technology*, 5(1), 1–10. <https://doi.org/10.30598/ppcst.knmxxiiv5i1p1-10>
- Geschwind, G., Vignal, M., Caballero, M. D., & Lewandowski, H. J. (2024). Evidence for validity and reliability of a research-based assessment instrument on measurement uncertainty. *Physical Review Physics Education Research*, 20(2), 020125. <https://doi.org/10.1103/PhysRevPhysEducRes.20.020125>
- Hair, J. F., Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate data analysis (8th ed.). *Cengage Learning EMEA*. <https://doi.org/10.5117/2006.019.003.007>

- Hakiki, A. W., Fitri, A. R., & Agung, I. M. (2018). Analisis properti psikometri subtes Merkaufgaben (ME) dengan Rasch model. *Jurnal Psikologi*, 14(1), 40–50. <https://doi.org/10.24014/jp.v14i1.4900>
- Haryanti, Y. D., & Saputra, D. S. (2019). Instrumen penilaian berpikir kreatif pada pendidikan abad 21. *Jurnal Cakrawala Pendas*, 5(2), 58–64. <https://doi.org/10.31949/jcp.v5i2.1350>
- Hidayah, N. C., Ulya, H., Masfuah, S., Guru, P., & Dasar, S. (2021). Analisis kemampuan berpikir kreatif siswa sekolah dasar berdasarkan tingkat kemampuan matematis. *Educatio*, 7(4), 1368–1377. <https://doi.org/10.31949/educatio.v7i4.1366>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Jonsson, B., Granberg, C., & Lithner, J. (2020). Gaining mathematical understanding: The effects of creative mathematical reasoning and cognitive proficiency. *Frontiers in Psychology*, 11, 574366. <https://doi.org/10.3389/fpsyg.2020.574366>
- Kavenuke, P. S., Kinyota, M., & Joel, J. K. (2020). The critical thinking skills of prospective teachers: Investigating their systematicity, self-confidence, and scepticism. *Educational Research Review*, 37, 100130.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Korobko, A. I. (2019). Measurement uncertainty as a test model assessment tool. 2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL) (pp. 1–4). *IEEE*. <https://doi.org/10.1109/CAOL46282.2019.9019418>
- Meika, I., & Sujana, A. (2017). Kemampuan berpikir kreatif dan pemecahan masalah matematis siswa SMA. *Jurnal Penelitian dan Pembelajaran Matematika*, 10(2), 8–13. <https://doi.org/10.30870/jppm.v10i2.2025>
- Morell, L., & Wilson, M. (2016). Assessment as a tool to understand students' conceptions of the structure of matter. *Journal of Physics: Conference Series*, 772(1). <https://doi.org/10.1088/1742-6596/772/1/012049>
- Mu'minah, I. H. (2021). Studi literatur: Pembelajaran abad-21 melalui pendekatan STEAM (Science, Technology, Engineering, Art, and Mathematics) dalam menyongsong era society 5.0. *Prosiding Seminar Nasional Pendidikan*, 3, 584–594.

- Nakano, T. de C., & Wechsler, S. M. (2018). Creativity and innovation: Skills for the 21st century. *Estudos de Psicologia (Campinas)*, 35(3), 237–246. <https://doi.org/10.1590/1982-02752018000300002>
- Narimawati, U., Sarwono, J., Affandi, A., & Priadana, S. (2020). Ragam analisis dalam metode penelitian. ANDI.
- Putra, H. D., Akhdiyati, A. M., Setiany, E. P., & Andiarani, M. (2018). Kemampuan berpikir kreatif matematik siswa SMP di Cimahi. *Jurnal Pendidikan dan Pembelajaran Matematika*, 9(1), 47–53.
- Putra, Y. (2020). Kemampuan berpikir kreatif siswa dalam pembelajaran matematika berbasis masalah. *Jurnal Pendidikan dan Pembelajaran Matematika*, 9(2), 1042–1045.
- Retnawati, H. (2017). Validitas, reliabilitas, dan karakteristik butir. Parama Publishing.
- Slamet, W., & Wardani, R. K. (2020). Mengajarkan keterampilan abad 21 4C (communication, collaboration, critical thinking and problem solving, creativity and innovation). *Jurnal Pendidikan Inovatif*, 7(3), 185–197.
- Sugiyono. (2019). Statistika untuk penelitian. Alfabeta.
- Sullivan, G. M. (2011). A primer on the validity of assessment instruments. *Journal of Graduate Medical Education*, 3(2), 119–120. <https://doi.org/10.4300/JGME-D-11-00075.1>
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial*. Trim Komunikata Publishing House.
- Susanti, E., Andrianti, S. D., Agung, D., & Kususmo. (2023). Peningkatan kemampuan berpikir kreatif siswa kelas V dengan pendekatan mind mapping. *Jurnal Pendidikan Dasar*, 9(1), 2416–2425.
- T, R., Laurens, T., & Moma, L. (2020). Analisis kemampuan berpikir kreatif matematis siswa SD Negeri 40 Ambon pada materi bangun datar. *JUMADIKA: Jurnal Magister Pendidikan Matematika*, 1(2), 91–101. <https://doi.org/10.30598/jumadikavol1iss2year2019page91-101>
- Tumangger, M. H., Kartono, & Ridlo, S. (2022). Analisis kebutuhan pengembangan instrumen penilaian kemampuan berpikir kreatif pada mata pelajaran matematika untuk siswa sekolah dasar. *Jurnal Basicedu*, 6(5), 9170–9179. <https://doi.org/10.31004/basicedu.v6i5.3623>
- Umar, A., & Ahmad, N. Q. (2019). Analisis kemampuan berpikir kreatif matematis calon guru matematika. *As-Salam: Jurnal Pendidikan dan Pembelajaran*, 3(1), 36–47.