Jurnal Manajemen Pendidikan *Volume 14, No. 01, 2026, Halaman 1241 – 1248*
https://jmp-pasca.unpak.ac.id/index.php/jmp *e-SSN: 2614-3313 ; p-ISSN: 2302-0296*
Penerbit: Sekolah Pascasarjana, Universitas Pakuan

DEVELOPMENT OF A SOLO TAXONOMY-BASED PROBLEM-SOLVING TASK INSTRUMENT IN TRIGONOMETRY OF FOSTER STUDENTS' CREATIVE THINKING

Crisnina Intan Setyo Maharani [a*)], Raden Sulaiman [a)], Agung Lukito [a)]

[a)] *Universitas Negeri Surabaya, Surabaya, Indonesia*

[*)]*e-mail korespondensi: 25030785007@mhs.unesa.ac.id*

Article history: received 06 Juni 2026; revised 20 Juni 2026; accepted 26 Juli 2026 *DOI : https://doi.org/10.33751/jmp.v14i01.292*

Abstrak. Berpikir kreatif sebagai salah satu kemampuan penting dalam pembelajaran matematika dapat diungkap melalui instrumen tugas pemecahan masalah yang berperan sebagai sarana memperoleh informasi mengenai proses berpikir siswa. Tujuan penelitian ini adalah untuk menghasilkan instrumen berupa tugas pemecahan masalah untuk melatih berpikir kreatif yang valid dan reliabel. Adapun subjek penelitian adalah 30 siswa kelas XI-4 SMAN 1 Gedangan tahun ajaran 2026/2027. Penelitian ini merupakan penelitian pengembangan dengan menggunakan model 4-D yaitu define, design, develop, dan disseminate. Terdapat 2 instrumen tugas pemecahan masalah yang dikembangkan masing-masing 1 butir soal dengan 5 subsoal yang disusun berdasarkan level Taksonomi SOLO. Validitas isi dilakukan oleh 3 validator mendapat kriteria validitas sangat tinggi, sedangkan validitas empirik diolah menggunakan korelasi Product Moment dan mendapat hasil validitas sedang hingga sangat tinggi. Hasil uji reliabilitas diperoleh koefisien reliabilitas masing-masing adalah 0.79 dan 0.84 dinyatakan reliabel atau memiliki reliabilitas cukup dan baik. Hasil tersebut menunjukkan bahwa instrumen tugas pemecahan yang dikembangkan layak untuk digunakan.

Kata Kunci: Instrumen Tugas Pemecahan Masalah, Taksonomi SOLO, Berpikir Kreatif

DEVELOPMENT OF A SOLO TAXONOMY-BASED PROBLEM-SOLVING TASK INSTRUMENT IN TRIGONOMETRY OF FOSTER STUDENTS' CREATIVE THINKING

Abstract. *Creative thinking, as one of the key competencies in mathematics learning, can be assessed through problem-solving task instruments, which serve as a means of obtaining information about students' thinking processes. This study aims to develop a valid and reliable problem-solving task instrument to foster creative thinking. The research subjects were 30 eleventh-grade students in Class XI-4 at SMAN 1 Gedangan during the 2026/2027 academic year. This study was a development research project using the 4-D model consisting of define, design, develop, and disseminate. There were two problem-solving task instruments developed, with one main question and five sub-questions each, structured according to the SOLO taxonomy levels. The content validity was assessed by three validators and received a "very high" validity rating, while the empirical validity was analyzed using the Product Moment correlation and yielded results ranging from "moderate" to "very high". The reliability test results showed reliability coefficients of 0.79 and 0.84, respectively, indicating that the instruments had "acceptable" and "good" reliability. These results demonstrate that the problem-solving task instruments developed are suitable for use*

Keywords: *Problem-Solving Task Instrument, SOLO Taxonomy, Creative Thinking*

I. INTRODUCTION

Mathematics is a discipline that emphasizes not only mastery of concepts and procedures but also the ability to apply that knowledge to solve various problems. In the context of learning, problem-solving is viewed as the core of mathematical activity as it involves the processes of understanding situations, making connections between various concepts, selecting appropriate strategies, and evaluating the solutions obtained (NCTM, 2000; Polya, 1973). Through these activities, students construct meaning around the concepts they are learning while demonstrating how mathematical knowledge is applied in non-routine situations. The variety of strategies and solution approaches that emerge during the problem-solving process makes this activity a rich source of

information about students' thinking processes, therefore necessitating an instrument capable of systematically capturing such information (Sapri, 2023).

A problem-solving task instrument serves as a means of obtaining information about students' thinking processes. Unlike routine questions, which are generally oriented toward the accuracy of the final answer, a problem-solving task instrument is designed to encourage students to articulate their reasoning, present representations, and select the strategies they consider most appropriate for solving a problem (Annizar et al., 2025). These characteristics allow teachers and researchers to gain a more comprehensive picture of students' abilities than mere calculation results. The information generated depends on the quality of the instrument used. Therefore, the development of such instruments requires a clear theoretical foundation and empirical evidence that the instrument is truly capable of assessing the construct being measured (Quezada & Albornoz, 2023).

Creative thinking is one of the abilities frequently studied in mathematics learning. This ability is reflected in students' proficiency in generating various alternative solutions (fluency), using a variety of strategies (flexibility), and discovering relatively new ideas or solutions (novelty) when faced with a problem (Silver, 1997). These characteristics are difficult to observe when students are presented only with procedural problems or those with a single solution pattern. In contrast, tasks that allow for the exploration of various possible solutions enable creative thinking skills to emerge more naturally. The close relationship between task characteristics and student responses indicates that the quality of task instruments is a crucial factor in fostering creative thinking more effectively.

Various studies have developed instruments to measure creative thinking abilities, but most are still dominated by tasks oriented toward the final product or that merely assess the accuracy of problem-solving procedures. As a result, the variety of strategies, arguments, and the complexity of students' thinking processes have not been fully documented. On the other hand, some instrument development studies have utilized a problem-solving approach, but these have not been entirely based on a framework capable of differentiating the quality of students' responses at various levels (Ulya et al., 2024). Consequently, the available instruments still have limitations in providing in-depth information regarding the development of students' creative thinking abilities across different response levels. These limitations necessitate the development of an instrument that considers both the characteristics of the material and the characteristics of the abilities to be assessed.

The selection of learning materials also plays a crucial role in the development of problem-solving task instruments, as the characteristics of each material influence the range of responses students can demonstrate. Trigonometry is one subject area with such potential because it involves interconnections between concepts, geometric representations, algebraic representations, and the application of concepts in various contextual situations (Juandi et al., 2025). Solving a trigonometry problem can often be approached through more than one strategy, whether by utilizing angle relationships, trigonometric identities, or different geometric representations. This diversity of approaches provides opportunities for students to demonstrate a variety of ideas, strategies, and reasoning used during the problem-solving process (Silver, 1997). It is this characteristic that makes trigonometry a relevant context for developing problem-solving task instruments aimed at fostering students' creative thinking abilities through the quality of the responses they produce.

The diverse nature of student responses requires a framework capable of organizing the quality of answers based on their level of complexity. One framework commonly used in mathematics learning research is the SOLO (Structure of the Observed Learning Outcomes) Taxonomy, developed by Biggs and Collis. The SOLO Taxonomy classifies student responses into five levels of development, which are prestructural, unistructural, multistructural, relational, and extended abstract, so that it not only assesses the accuracy of answers but also indicates the quality of understanding reflected in the structure of students' responses (Biggs & Collis, 1982). In the context of developing problem-solving task instruments, this framework provides a systematic basis for constructing test items with varying levels of complexity while helping to identify students' abilities to develop and connect various ideas when solving mathematical problems (Bicknell & Riley, 2012).

The use of the SOLO Taxonomy in mathematics learning research has expanded across various contexts, both as a framework for analyzing the quality of student responses and as a basis for developing learning instruments (Maryanti et al., 2018). This research indicates that the SOLO Taxonomy is capable of distinguishing levels of student understanding more comprehensively than assessments that focus solely on whether answers are correct or incorrect. Meanwhile, research on instruments for fostering creative thinking abilities continues to advance through the use of problem-solving tasks and open-ended questions. However, these two approaches still tend to develop separately. The development of creative thinking instruments generally emphasizes indicators of creativity, whereas the SOLO Taxonomy is more frequently used to identify students' levels of understanding or the quality of their responses.

These conditions indicate an opportunity to integrate the characteristics of problem-solving tasks, indicators of creative thinking ability, and the level of response complexity into a single, comprehensive instrument. This integration is essential to ensure that the instrument not only elicits various alternative solutions but also provides information on whether the quality of students' responses reflects creative thinking (Febrima & Mahmudi, 2026). This study developed a problem-solving task instrument for trigonometry by utilizing the SOLO Taxonomy as the basis for constructing test items, and then tested the quality of the instrument through content validity, empirical validity, and reliability. This focus distinguishes this study from previous research, which generally emphasized the validation of the instrument's content or the use of the SOLO Taxonomy as a tool for analyzing student responses.

This study aims to develop a SOLO Taxonomy-based problem-solving task instrument for trigonometry to foster students' creative thinking abilities that meet validity and reliability criteria. The resulting instrument is expected to serve as an alternative for teachers and researchers in obtaining information about students' creative thinking abilities through problem-solving tasks.

II. METHOD

This study was a research and development (R&D) study aimed at producing a SOLO Taxonomy-based problem-solving task instrument to foster students' creative thinking. The development model used was the 4-D model developed by Thiagarajan et al. (1974), which included the define, design, develop, and disseminate stages.

The define stage aimed to identify the needs for instrument development. Activities in this stage included an analysis of trigonometry material as well as a review of theories regarding creative thinking and the SOLO Taxonomy. The results of the analysis were used as the basis for determining the specifications of the instrument to be developed.

The design stage aimed to develop a preliminary draft of the instrument. The activities included the development of an instrument blueprint, the creation of two problem-solving tasks, scoring guidelines, and a validation sheet. The instrument was designed in accordance with the conceptual material on trigonometric ratios and the characteristics of each level of the SOLO taxonomy, including unistructural, multistructural, relational, and extended abstract.

The develop stage aimed to produce an instrument that met the criteria for validity and reliability. The instrument's validity was assessed by two mathematics education lecturers and one mathematics teacher through an evaluation of the content, construct, and language aspects. Feedback from the validators was used as the basis for revising the instrument until a suitable version was obtained. The revised instrument was then administered to 30 eleventh-grade students in Class XI-4 at SMAN 1 Gedangan to obtain empirical data on the instrument's validity and reliability.

Content validity was analyzed quantitatively by calculating the average score for each assessment aspect using Aiken's (1985) formula as follows.

$$V = \frac{\sum s}{n(c-1)}$$

Description:
V: item validity index
s : r-lo
$\sum s$: $s_1 + s_2$ + etc.
n : number of raters
c : highest validity rating score (e.g., 5)
lo : lowest validity rating score (e.g., 1)
r : rating given by the validator

Table 1. Content Validity Criteria

Value	Criteria
$0.80 < V \leq 1.00$	Very High
$0.60 < V \leq 0.80$	High
$0.40 < V \leq 0.60$	Moderate
$0.20 < V \leq 0.40$	Low
$0.00 < V \leq 0.20$	Very Low

After the instrument was determined as valid in terms of content based on assessments by experts and practitioners, empirical validity testing was conducted using Pearson's Product Moment correlation by correlating the score of each item with the total score (Sari et al., 2023).

$$r_{xy} = \frac{N \sum xy - \sum x \sum y}{\sqrt{(N\sum x^2 - (\sum x)^2)(N\sum y^2 - (\sum y)^2)}}$$

Description:
r_{xy}: validity coefficient
N: number of subjects
x : score of the item for which validity is being assessed
y : total score

Table 2. Empirical Validity Criteria

Correlation Coefficient	Validity Criteria
$0.80 < r_{xy} \leq 1.00$	Very High Validity
$0.60 < r_{xy} \leq 0.80$	High Validity
$0.40 < r_{xy} \leq 0.60$	Moderate Validity
$0.20 < r_{xy} \leq 0.40$	Low Validity
$0.00 < r_{xy} \leq 0.20$	Very Low Validity
$r_{xy} = 0$	Invalid

The instrument's reliability was analyzed using Cronbach's Alpha to determine its level of internal consistency (Revita et al., 2018). The instrument was considered reliable if it had a Cronbach's Alpha coefficient of ≥ 0.70. indicating acceptable internal consistency (Taber, 2018).

$$\alpha = \frac{k}{k-1}\left(1-\frac{\sum S_i^2}{S_t^2}\right)$$

Description:
α : reliability coefficient (Cronbach's Alpha)
k : number of instrument items
$\sum S_i^2$: sum of the variances of each item
S_t^2 : total variance of the instrument score

Table 3. Reliability Criteria

Cronbach's Alpha Value	Reliability Criteria
$\alpha \geq 0.90$	Excellent
$0.80 \leq \alpha < 0.90$	Good
$0.70 \leq \alpha < 0.80$	Acceptable
$0.60 \leq \alpha < 0.70$	Questionable
$0.50 \leq \alpha < 0.60$	Poor
$\alpha < 0.50$	Unacceptable

Furthermore, the dissemination stage was conducted on a limited basis by introducing the developed instrument to mathematics teachers at the schools where the instrument was tested. The purpose of this stage was to ensure that the developed instrument could be utilized as a tool to foster students' creative thinking in mathematics learning.

III. RESULT AND DISCUSSION

Content Validity

Content validity is the initial step in ensuring that each item in the instrument represents the construct to be measured. The results of the content validity are presented in Table 4 and Table 5.

Table 4. Validity Results of Problem-Solving Task Instrument 1

Aspect	Item Number	Validator			V	Criteria
		V1	V2	V3		
A1	1a	4	5	5	0.91	Very high
A2		5	5	5	1	Very high
A3		5	5	5	1	Very high
A4		5	5	5	1	Very high
A1	1b	4.7	5	5	0.97	Very high
A2		5	5	5	1	Very high
A3		5	5	5	1	Very high
A4		5	5	5	1	Very high

A1		5	5	5	1	Very high
A2	1c	5	5	5	1	Very high
A3		5	5	5	1	Very high
A4		5	4.7	5	0.97	Very high
A1		4.3	5	5	0.94	Very high
A2	1d	5	5	5	1	Very high
A3		5	5	5	1	Very high
A4		5	4	5	0.91	Very high
A1		5	5	5	1	Very high
A2	1e	4.3	5	5	0.94	Very high
A3		5	5	5	1	Very high
A4		4.3	4.7	4.3	0.85	Very high

Based on the validation results conducted by three validators consisting of two mathematics education lecturers and one mathematics teacher, the problem-solving task 1 instrument was evaluated based on four aspects, including material appropriateness (A1), construct (A2), complexity (A3), and language (A4) for each item. The validation results showed that the instrument obtained validity coefficients (V) ranging from 0.85 to 1.00 with all aspects categorized as "very high". Most aspects received a validity score of 1.00. while the lowest score of 0.85 was found in the language aspect (A4) of item 1e. These results indicate that each test item met the content validity criteria for every aspect and is therefore suitable for use in the testing stage. Although the validators provided several suggestions for improvement, these suggestions were limited to refinements in the wording and presentation of the items without altering the substance of the instrument. Consequently, the problem-solving task 1 instrument is considered valid and can be used in research after revisions were made based on the validators' feedback.

Table 5. Validity Results of Problem-Solving Task Instrument 2

Aspect	Item Number	Validator			V	Criteria
		V1	V2	V3		
A1		4	5	5	0.91	Very high
A2	1a	5	5	5	1	Very high
A3		5	5	5	1	Very high
A4		5	5	5	1	Very high
A1		4.7	5	5	0.97	Very high
A2	1b	5	5	5	1	Very high
A3		5	5	5	1	Very high
A4		5	5	5	1	Very high
A1		5	5	5	1	Very high
A2	1c	5	5	5	1	Very high
A3		5	5	5	1	Very high
A4		5	4.7	5	0.97	Very high
A1		4.3	5	5	0.94	Very high
A2	1d	5	5	5	1	Very high
A3		5	5	5	1	Very high
A4		5	4	4.3	0.85	Very high
A1		5	5	5	1	Very high
A2	1e	4.3	5	5	0.94	Very high
A3		5	5	5	1	Very high
A4		4.3	4.7	4.3	0.85	Very high

The problem-solving task 2 instrument based on Table 5, obtained validity coefficients (V) ranging from 0.85 to 1.00 with all test items categorized as "very high." Most assessment components achieved a validity score of 1.00. while the lowest score of 0.85 was found in test items 1d and 1e. Overall, these results indicate that the Problem-Solving Task 2 instrument has met the criteria for content validity and is considered suitable for use in the testing stage. Several suggestions provided by the validators were used as a basis for refining the instrument, thereby improving its quality without altering the substance or the established measurement objectives.

The content validity results showed that the developed instrument was considered appropriate by the validators based on the aspects of content appropriateness, construct, complexity, and language. These findings indicate that each item in the instrument met the necessary criteria for use in the testing stage after refinements were made based on the validators' feedback.

These findings are consistent with the research by Kania et al. (2024), which reported that instruments that achieved very high content validity through expert assessment were considered suitable to proceed to the empirical testing stage after revisions were made to editorial aspects. Similar results were also found by Nurjanah et al. (2023), who demonstrated that expert involvement in evaluating the appropriateness of content, language, and presentation can improve the quality of the instrument without altering the substance of the developed items. The similarity of these findings indicates that the expert validation process plays a crucial role in ensuring the instrument's suitability before it is used to collect research data.

Empirical Validity

Empirical validity was assessed to determine whether each item is valid as a data collection instrument. The results of the empirical validity tests for problem-solving task 1 and problem-solving task 2 are presented in Table 6 and Table 7.

Table 6. Empirical Validity Results of Problem-Solving Task 1

Item	$r_{calculated}$	Criteria
1a	0.63	High Validity
1b	0.46	Moderate Validity
1c	0.89	Very High Validity
1d	0.77	High Validity
1e	0.79	High Validity

The results of the empirical validity test for problem-solving task 1 showed that all test items had a positive correlation coefficient with the total score, so all items were considered valid. This indicates that each item is able to contribute to measuring the same ability as the instrument as a whole. The levels of validity among the items ranged from moderate to very high, indicating that each item's representativeness of the total score was not entirely consistent. Therefore, all items met the validity criteria and remain suitable for use in the data collection stage. These findings show that the problem-solving task 1 instrument possesses adequate item quality to assess students' creative thinking abilities.

Table 7. Empirical Validity Results of Problem-Solving Task 2

Item	$r_{calculated}$	Criteria
1a	0.63	High Validity
1b	0.56	Moderate Validity
1c	0.95	Very High Validity
1d	0.88	Very High Validity
1e	0.90	Very High Validity

The results of the empirical validity test for problem-solving task 2 revealed that the quality of the test items tended to be better than that of problem-solving task 1. Most items were categorized as very high in validity, with one item having a moderate validity rating while still meeting the validity criteria. These results indicate that each item has a strong relationship with the total score, thereby providing consistent information regarding the ability being measured. The improvement in item quality in problem-solving task 2 also demonstrates that the revision process based on logical validation and instrument testing had a positive impact on the quality of the instrument, particularly in enhancing the ability of each item to distinguish variations in student ability.

The findings of this study are consistent with the research by Mahmuzah & Aklimawati (2022), which developed a mathematical disposition scale and reported that items with a Product Moment correlation met the validity criteria and were therefore retained in the final instrument. These results indicate that an adequate correlation between item scores and total scores serves as evidence that each item contributes to the overall measurement. Furthermore, the results of this study are consistent with the research by Lusiana et al. (2024), which demonstrated that instruments tested using Pearson product-moment correlations yielded valid items, making them suitable for use as research instruments after meeting the criteria for empirical validity. Thus, the empirical validity obtained for both problem-solving tasks provides evidence that all test items are suitable for use in training students' creative thinking abilities.

Reliability

After all test items were empirically validated, the next step was to test the instrument's reliability to determine the level of consistency in the measurement results, as presented in Table 8 and Table 9.

Table 8. Reliability Results of Problem-Solving Task 1

Reliability Coefficient	Criteria
0.79	Acceptable

Based on Table 8, the problem-solving task 1 instrument obtained a reliability coefficient of 0.79, which fell into the "acceptable" reliability category. These results indicate that the instrument has an adequate level of consistency in measuring mathematical creative thinking abilities and can therefore be used as a research instrument. A reliability value approaching the high category also indicates that each test item possesses good internal consistency in generating measurement data.

Table 9. Reliability Results of Problem-Solving Task 2

Reliability Coefficient	Criteria
0.84	Good

The problem-solving task 2 instrument achieved a reliability coefficient of 0.84, which fell into the high reliability "good" category. These results indicate that the instrument has a higher level of internal consistency than problem-solving task 1. Therefore, the problem-solving task 2 instrument is capable of providing relatively stable measurement results when used to assess students' mathematical creative thinking abilities.

The research results showed that both instruments had reliability levels that met the criteria for use in research. The reliability coefficient for problem-solving task 1 indicated that the instrument had adequate internal consistency, while problem-solving task 2 demonstrated higher internal consistency. This difference in reliability coefficients indicates that the quality of inter-item correlations in problem-solving task 2 is better, enabling it to produce more consistent measurements. These findings suggest that the instrument development process, including expert validation, item revision, and instrument testing, contributed to the formation of an instrument with a high level of measurement consistency.

The findings of this study are consistent with the research by Rahayu et al. (2025), which showed that Cronbach's Alpha is used to indicate the internal consistency of an instrument, where a higher value suggests that the test items are more consistent in measuring the same ability. The results of this study are also consistent with the research by Taber (2018), which reported that instruments with an Alpha coefficient above 0.70 have sufficient reliability for use in educational research, while higher values indicate better internal consistency. Therefore, the reliability obtained for all problem-solving tasks provides evidence that both instruments are capable of producing consistent data, making them suitable for use in developing students' creative thinking abilities.

IV. CONCLUSION

This study successfully developed a SOLO Taxonomy-based problem-solving task instrument to foster students' creative thinking abilities in mathematics, focusing on trigonometry. The results indicate that the instrument met the criteria for high quality supported by quantitative evidence. The content validity of both Problem-Solving Task 1 and Problem-Solving Task 2 was categorized as "very high" based on expert evaluation, with Aiken's V indices ranging from 0.85 to 1.00 across all aspects. Furthermore, empirical validity indicated that all test items met the validity criteria based on their correlation with total scores, yielding r_calculated values ranging from 0.46 to 0.95. The reliability test results indicated that the instruments were reliable, with Problem-Solving Task 1 exhibiting internal consistency in the "acceptable" category and Problem-Solving Task 2 in the "good" category. Therefore, the developed instruments meet the requirements as valid and reliable measurement tools for fostering students' creative thinking abilities.

The developed SOLO Taxonomy-based instrument provides an alternative assessment instrument that can foster creative thinking abilities through problem-solving tasks while considering the complexity level of students' responses. Crucially, the successful validation demonstrates that employing the SOLO Taxonomy specifically as a framework for designing and structuring the test items effectively accommodates the varying complexities of students' responses to foster their creative thinking. It is hoped that this instrument can be utilized by researchers and teachers as a reference in developing instruments that align with the characteristics of the subject matter and the abilities being measured. Future research is recommended to test the instrument across a broader range of subjects and other mathematics topics to obtain more comprehensive evidence of its validity and reliability.

V. REFERENCES

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. Educational and Psychological Measurement, 45(1), 131–142. https://doi.org/10.1177/0013164485451012

Annizar, A. M., Kurniati, D., & Yudianto, E. (2025). Towards a comprehensive framework for complex problem-solving in mathematics: A development and validation study. TEM Journal, 14(3), 1–11. https://doi.org/10.18421/tem143-85

Bicknell, B., & Riley, T. (2012). Assessing students' mathematical problem-solving abilities using the SOLO taxonomy (J. Dindyal, L. P. Cheng, & S. F. Ng (eds.); Mathematic). MERGA.

Biggs, J., & Collis, K. F. (1982). Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome). Academic Press. https://doi.org/10.1177/000494418302700311

Febrima, R., & Mahmudi, A. (2026). Analysis of mathematical creative thinking ability based on Solo's Taxonomy. Jurnal Eduscience (JES), 13(3), 999–1012. https://doi.org/10.36987/jes.v13i3.9315

Juandi, Sugiatno, & Noviani, E. (2025). Enhancing students' conceptual understanding of trigonometric functions through geogebra based-learning. Al-Jabar: Jurnal Pendidikan Matematika, 16(01), 55–70. https://doi.org/10.24042/ajpm.v16i1.25653

Kania, N., Kusumah, Y. S., Dahlan, J. A., Nurlaelah, E., & Gürbüz, F. (2024). Constructing and providing content validity evidence through the Aiken ' s V index based on the experts ' judgments of the instrument to measure mathematical problem-solving skills. Research and Evaluation in Education (REiD), 10(1), 64–79. https://doi.org/10.21831/reid.v10i1.71032

Lusiana, I. S., Susongko, P., & Nafiati, D. A. (2024). Desain instrumen tes capaian pembelajaran matematika dengan uji validitas Pearson Correlation. Journal of Education Research, 5(4), 5666–5675. https://doi.org//10.37985/jer.v5i4.1700

Mahmuzah, R., & Aklimawati, A. (2022). Pengembangan instrumen skala disposisi matematis. Jurnal Pendidikan Matematika Malikussaleh, 2(1), 229–238. https://doi.org/10.29103/jpmm.v2i1.7462

Maryanti, Y., Abdul, D., Lidinillah, M., & Rokhayati, A. (2018). Pengembangan model penilaian berbasis taksonomi solo pada konsep pembagian pecahan sekolah dasar. PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar, 1(2), 103–112.

NCTM. (2000). Principles and Standards for School Mathematics.

Nurjanah, S., Istiyono, E., Widihastuti, W., & Iqbal, M. (2023). The application of Aiken's V method for evaluating the content validity of instruments that measure the implementation of formative assessments. Journal of Research and Educational Research Evaluation, 12(2). https://doi.org/10.15294/jere.v12i2.76451

Polya, G. (1973). How To Solve It. Princeton University Press.

Quezada, V. D., & Albornoz, M. S. (2023). Validation of an instrument to assess deductive reasoning in solving types of problems. International Journal of Engineering Pedagogy, 13(7), 50–64. https://doi.org/10.3991/ijep.v13i7.40153

Rahayu, T. G., Nurussama, A., Putri, H. E., Salma, R., Khaerani, M., & Manalu, B. (2025). Pengembangan instrumen tes uraian yang valid dan reliabel untuk mengukur kemampuan berpikir kritis matematis siswa kelas V SD Guru Sekolah Dasar. Delta-Pi: Jurnal Matematika Dan Pendidikan Matematika, 14(2), 137–154.

Revita, R., Kurniati, A., & Andriani, L. (2018). Analisis instrumen tes akhir kemampuan komunikasi matematika untuk siswa smp pada materi fungsi dan relasi. Jurnal Cendekia: Jurnal Pendidikan Matematika, 2(2), 8–19.

Sapri, H. A. (2023). Analysis of mathematics problem solving ability based on Polya problem solving steps. International Journal of Research in Educational Review, 2(2), 177–187. https://doi.org/10.51574/ijrer.v2i2.755

Sari, D. P., Darhim, & Rosjanuardi, R. (2023). Pengembangan instrumen tes pemecahan masalah aljabar berbasis Taksonomi SOLO. Jurnal Cendekia: Jurnal Pendidikan Matematika, 7(2), 1538–1551. https://doi.org/10.31004/cendekia.v7i2.2265

Silver, E. A. (1997). Fostering Creativity through Instruction Rich in Mathematical Problem Solving and Problem Posing. ZDM – Mathematics Education, 29(3), 75–80.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. Research in Science Education, 48(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). Instructional development for training teachers of exceptional children. University of Minnesota.

Ulya, H., Sugiman, & Rosnawati, R. (2024). 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), 1–11. https://doi.org/10.29333/ejmste/14661