



Algebra essay questions: A validated instrument for evaluating the students' higher-order thinking skills

Nia Kania ^{1*}, Florence M. Kyaruzi ², Lilis Marina Angraini ³

¹ Mathematics Education, Universitas Majalengka, Indonesia

² Department of Educational Psychology and Curriculum Studies, Dar es Salaam, Tanzania

³ Mathematics Education, Universitas Islam Riau, Indonesia

* Corresponding Author. Email: niakania@unma.ac.id

Received: 6 August 2023; Revised: 16 August 2023; Accepted: 15 September 2023

Abstract: Indonesian pupils need help solving problems, particularly in mathematics, due to the need for practice questions with the higher-order thinking (HOT) type. This research aims to develop a reliable and valid tool for assessing students' higher-order thinking skills in mathematics. Developing a test involves gathering requirements, creating questions, assessing answers, running trials, analyzing data, making adjustments, and finally putting everything together. The instrument trial was conducted at SMAN at Majalengka, West Java. The total number of test subjects was 32 students. Research data were collected using a description test. The test instrument consists of 10 descriptions, but only 6 had good validity based on the value of V on the V-Aiken. In addition, the test package produces reliable questions based on Cronbach's Alpha scores. The test results of the instrument showed that the HOTS of mathematics for Grade X students was good. This step provides an overview to teachers in designing valid and relevant instruments.

Keywords: Algebra essay, Higher-order thinking skill, Validated instrument

How to cite: Kania, N., Kyaruzi, F. M., & Angraini, L. M. (2023). Algebra essay questions: A validated instrument for evaluating the students' higher-order thinking skills. *Union: Jurnal Ilmiah Pendidikan Matematika*, 11(3), 367-375. <https://doi.org/10.30738/union.v11i3.15688>

INTRODUCTION

Education worldwide has increasingly emphasized the cultivation of higher-order thinking skills (HOTS) among students (Priyanto & Dharin, 2021; Korkmaz et al., 2020), which will prepare them for the challenges of the fourth industrial revolution, or the digital era (Lyons & Bandura, 2020; Miri, 2007; Moussa et al., 2020). The advent of the fourth industrial revolution, often referred to as the digital era, has ushered in unprecedented technological advancements and transformative changes across various sectors.

HOTS are essential in Salihu et al. (2018) and Wang & Wang, (2011) of both teaching and learning (Hopson et al., 2001; Tanujaya et al., 2017). HOTS are significant components of both teaching (Salihu et al., 2018; Wang & Wang, 2011) and learning (Hopson et al., 2001; Tanujaya et al., 2017). In this period of change, the world has recognized that students must be able to develop higher-order thinking and make the right choices. HOTS encompass cognitive abilities beyond memory, requiring exceptional problem-solving skills to tackle novel challenges and ambiguities.

The educational system in Indonesia is highly centralized with a well-defined hierarchy of educational decision (Balzano, 1991). The implementation of HOTS-based curricula (Setiawan et al., 2021) as described in the "standard content" for mathematics education, is one of the government's initiatives (Ozmantar, 2017; Tanudjaya & Doorman, 2020). According to KTSP



and the current curriculum (2013), learning mathematics in Indonesia aims to foster students' HOTS, include teamwork, communication, critical thinking, and creativity. Thus, the Indonesian government has been working to create HOTS for math students. The HOTS-based assessment tool has now been integrated into the National Examination Assessment (UN) (Ichsan et al., 2020). The incorporation of HOTS-based assessment into the National Examination (UN) is fundamentally transforming students' educational trajectories and exerting a profound impact on their academic accomplishments. This integration is fostering the development of critical thinking, problem-solving prowess, and comprehensive cognitive capacities among students. Moreover, it is imperative to delve into the specific mechanisms through which this integration augments these skills. The Indonesian educational landscape presents a variety of needs and challenges that require careful consideration and innovative solutions.

The Programme for International Student Assessment (PISA) is an international assessment conducted by the Organization for Economic Co-operation and Development (OECD) that assesses the skills and knowledge of 15-year-old students in various countries. PISA aims to assess how well students can apply what they have learned in real-life situations, not just memorize tests. The measures students' critical thinking skills and problem-solving ability in science and mathematics at the age of fifteen and measures higher-order thinking skills (OECD, 2016). PISA is designs a mathematics literacy test for 15-year-old learners (Tasman, 2020). OECD also emphasizes the significance of students acquiring information literacy skills as part of their education (Kania & Juandi, 2023; Wijaya, 2016) These skills are part of higher-order thinking (HOT) skills (Tanujaya et al., 2017). According to the findings of various foreign polls, Indonesian pupils struggle to solve problems, particularly in mathematics. This difficulty could mainly be attributed to the lack of students practicing questions with the HOTS type.

The shortage of practice Higher Order Thinking Skills (HOTS) type questions among Indonesian students is a concerning issue that can have significant implications for their educational development. Thus, this research can contribute to developing questions to improve students' HOTS in the algebraic discussion so that these results can provide an overview of the stages of developing a HOTS measuring instrument. In addition, the more HOTS questions created will provide a reference for students to practice HOTS type questions. In terms of its operational goals, this study sought to: (1) develop HOTS indicators for senior high school mathematics; (2) verify the validity of the HOTS instrument; and (3) ascertain the reliability of the HOTS instrument.

The research focuses on validating algebra essay questions as a reliable tool for assessing students' higher-order thinking skills. Based on some of the models above, this study formulated aspects of critical thinking skills While higher-order thinking skills are widely recognized as important in education, there is a paucity of research that systematically validates the effectiveness of algebraic essay questions in assessing these skills. Previous studies may have touched on the concept, but comprehensive instrument validation is lacking.

METHOD

The research and development (R&D) approach is a systematic and iterative process that involves investigating, designing, and creating innovative solutions to address specific challenges or to enhance existing methodologies, products, or systems. The selection of the R&D approach was driven by its capacity to provide a structured framework for not only exploring novel avenues but also for refining and optimizing existing strategies. This method was chosen due to its inherent flexibility and adaptability, allowing for continuous refinement and adjustment based on emerging insights. Its alignment with the study's goals is rooted in the pursuit of generating practical, impactful, and sustainable outcomes that directly cater to the identified research objectives.

This research yielded a product in the form of an instrument that teachers can use as an example or reference to measure students' HOTS in mathematics. The research framework

utilized in this study is a derivative of the development model proposed by [Mardapi \(2008\)](#). This model encompasses a series of sequential steps, including: (1) gathering test specifications, (2) formulating test inquiries, (3) investigating test queries, (4) conducting preliminary test runs, (5) scrutinizing the items, (6) refining the test based on findings, (7) constructing the final test format, (8) conducting the test administration, and (9) interpreting the resultant test outcomes. In this study, however, the researcher restricted the development model to assembling the test, to apply the test, further research is needed in other words, the researcher did not conduct the test interpret the results because the last two steps might be carried out by teachers who are also employed in teaching in schools. This instrument was tested on 32 students at SMA 2 Majalengka, Majalengka Regency, 12 male and 20 female students. This instrument was tested on 32 grade XI high school students in Majalengka Regency, Indonesia. The total population who took part in the study was determined through the use of a method called purposeful sampling, the researcher opted to select an existing class for the study.

This research used of the following approaches to data analysis: content validity, construct validity, reliability, level of difficulty, and discrimination, and students' mathematical HOTS abilities. The content validity technique involved asking four experts in mathematics education, as validators, to test the accuracy and provide an assessment the suitability of the items and their indicators, and the editor of the composition of the questions.

Following the expert's assessment, the researcher calculated the outcomes of the assessment using a validity index, incorporating the Aiken index in [Formula \(1\)](#).

$$V = \frac{\sum s}{N(c-1)}, s = r - l \dots\dots\dots (1)$$

[\(Aiken, 1980\)](#)

In this context, the provided information pertains to the variables involved in the evaluation process. The possible range of numbers, denoted as V, represents the spectrum within which assessments or ratings can fall. The variable "r" corresponds to the appraiser's rating, indicating the evaluation given by the individual responsible for assessing a particular subject. On the other hand, "l" signifies the lowest category rater rating, signifying the assessment from the least favorable rater or evaluator. Conversely, "c" denotes the highest category rating, which reflects the evaluation from the most favorable rater. The variable "N" stands for the total number of raters or respondents participating in the evaluation process. These variables collectively contribute to the comprehensive understanding of the appraisal and rating procedures being employed.

The range of numbers is 0 to 1. While the validity of an item is higher, the number V will be closer to or equal to 1, and when the validity of an item is lower, the number V will be closer to or equal to 0. Similarly, the range of products available on the market has shrunk ([Aiken, 1980](#)). If the index is less than or equal to 0.4, then it is considered to have a moderate level of validity; if it is between 0.4 and 0.8, then it is considered to have a high level of validity; and if it is more significant than 0.8, then it is considered to have high level of validity ([Retnawati, 2014](#)).

In addition, exploratory component analysis was applied to demonstrate the construct's validity. The percentage of variance seen from the Kaiser Meyer Olkin (KMO) result indicates exploratory factor analysis ([Taştan & Yilmaz, 2008](#)). The IBM SPSS 20 Software was used to calculate the KMO value, and the sample used enables for additional investigation ([Santoso, 2006](#)).

During this time, an internal consistency method, the Cronbach-alpha formula, and IBM SPSS Software 20 were used to estimate of the instrument's reliability. Cronbach's Alpha values between 0.60 and less than 1 show that the instrument has satisfied the dependable criterion; on the other hand, if the Cronbach's Alpha value is less than 0.50, it indicates that the instrument is not reliable ([Basuki & Hariyanto, 2014](#); [Surapranata, 2009](#)). Meanwhile, according to [Brookhart and Nitko \(2011\)](#) the differentiating power of description questions is calculated using the [Formula \(2\)](#).

$$\text{distinguishing power} = \frac{\text{top group mean} - \text{bottom group mean}}{\text{maximum score of the question}} \dots\dots\dots (2)$$

After the calculation, the items are sorted into three categories: acceptable, revised, and rejected. It depends on the coefficient of difference. If there are questions that are rejected, they can be discarded or replaced with new items. Category for differential power coefficient is $DB > 0,3$ (received), $0,10 \leq DB < 0,30$ (revised), and $DB < 0,10$ (rejected) (Surapranata, 2009).

After the instrument met the criteria set out above, data analysis was carried out to determine the students' HOTS. The criteria used are if the results of the assessment show the student's mathematics HOTS score is more than or equal to 65 (on a scale of 0-100), then the student's mathematics HOTS has met the excellent criteria.

RESULTS AND DISCUSSION

The examination of the instrument, which was carried out by four highly qualified individuals in the field of mathematics, produced evidence demonstrating that the contents of the instrument are accurate. The four people are math lecturers with algebra skills, math education instrument experts, school mathematicians, and a teacher with experience in schools. The study results showed that the initial instrument made by the researcher could have been better. Therefore, the researcher improved/revised as much as possible according to the suggestions written by the validator on the instrument sheets. Following the conclusion of the maintenance, the instrument was returned to the validator so that it could be evaluated for each component. In order to evaluate whether or not each item on the test was valid, the outcomes of the assessment were analyzed using the Aiken formula, as presented in Table 1.

Table 1. Value V Aiken

No Question	Value V Aiken	Percentage (%)
1	0,56	16,67
2	0,44	10,00
3	0,28	3,33
4	0,50	5,00
5	0,17	10,00
6	0,56	16,67
7	0,22	3,33
8	0,67	11,67
9	0,28	5,00
10	0,61	18,33

Out of the initial set of 10 questions, a total of 4 questions were identified to have issues concerning their validity. Following consultations with high school teachers and considering time constraints, these 4 questions were subsequently removed from the instrument. As a result, the final assessment now comprises 6 items, carefully selected based on their medium and high levels of validity. With this revised instrument, the evaluation process can be efficiently conducted within a 90-minute timeframe, ensuring a more focused and reliable approach to gauging the desired outcomes.

After finishing the test questions in the field, the researcher carried out the tasks involved in scoring the test. Exploratory factor analysis was used to show that this is true about the construct. Table 2 presents the findings obtained from conducting the exploratory factor analysis. Examined the following results of the factor analysis.

Table 2. Explainable Variance

Explainable Variance	Category
51,2%	Accepted
5,7%	Revised
43,1%	Rejected

According to the findings presented in [Table 2](#), it is evident that over 50% of the variance can be explained. This substantial percentage indicates that the chosen variables and samples have contributed significantly to the analysis. The high level of explained variance provides researchers with a strong foundation to delve deeper into the data and conduct further analyses confidently. The ample variance allows for a comprehensive exploration of the relationships and patterns within the dataset, leading to more meaningful insights and a more comprehensive understanding of the phenomenon under study. With such promising results, researchers can proceed with additional investigations and confidently draw robust conclusions, thereby advancing knowledge and contributing to the field.

The assessment of the newly developed HOTS instrument demonstrated favorable outcomes. The reliability estimation of the instrument revealed a Cronbach's Alpha coefficient of 0.89239, indicating a high level of internal consistency and reliability within the test package. This robust coefficient underscores the dependability of the item package, suggesting that the HOTS instrument can be confidently utilized to measure and evaluate higher-order thinking skills effectively. Researchers and educators can place trust in the instrument's ability to provide consistent and accurate results, allowing for more reliable assessments of students' higher-order cognitive abilities. This finding adds credibility to the validity of the instrument, paving the way for its potential widespread application in educational settings to promote and enhance critical thinking and problem-solving skills.

The next thing that needs to be done in the activity that involves analysis is to figure out how challenging each question in the set is to find out the difficulty index, which is calculated using [Formula \(3\)](#).

$$\text{Difficulty Index} = \frac{M_{S(i)} + M_{W(i)}}{N \cdot m_i} \dots\dots\dots (3)$$

Where, $M_{S(i)}$ = the amount score of higher groups; $M_{W(i)}$ = the amount score of lower group; N = the respondents; and m = maximum score of each question. According to the following classification: very easy (.81 to 1.00), easy (.61 to .80), medium (.41 to .60), difficult (.21 to .40), and extremely difficult (.00 to .20) ([Desstya et al., 2019](#)). It is to see that the instrumentation of the device has varying degrees of difficulty; therefore, it can be said that the questions were heterogeneous. The results of the difficulty index can be seen in [Table 3](#).

Table 3. Difficulty Index

Question Number	Difficulty Index
1	0.52
2	0.51
3	0.45
4	0.61
5	0.29
6	0.27

The analysis of [Table 3](#) reveals that a majority of the questions were strategically designed to cover a range of difficulty levels, with a distribution spanning from medium to challenging. This careful arrangement of question difficulty levels was determined through a comprehensive study of the elements present in each question. By incorporating a mix of moderate to more complex questions, the assessment aims to effectively gauge students' proficiency across various skill levels. Such a diversified approach ensures that the evaluation process provides a comprehensive and accurate measure of students' abilities, capturing both their strengths and areas that require further development. This balanced difficulty distribution enhances the overall validity and reliability of the assessment, enabling educators and researchers to make more informed decisions and gain deeper insights into students' performance and progress.

The next thing that needs to be done in the analysis is to determine the distinguishing power of each item in each set of questions. The data analysis findings in [Table 4](#) demonstrate that

not all of the items had excellent discriminating power; some items that need to be amended, improved, or replaced, and need to be removed entirely.

Table 4. Power of Difference

Question Number	Difficulty Index
1	0.57
2	0.49
3	0.57
4	0.44
6	0.48
6	0.23

The insights gleaned from [Table 4](#) indicate that the items included in the question package demonstrate strong discriminating power. Notably, all the items were found to be reliable, and none were discarded during the analysis. This favorable outcome underscores the effectiveness of the selected items in effectively differentiating between high and low performers. The absence of item rejections further reinforces the instrument's integrity and accuracy in assessing students' abilities and aptitudes. As a result, educators and researchers can place confidence in the reliability of the question package, ensuring that it can yield precise and meaningful results. The robust discriminating power of the items enables educators to obtain a more comprehensive and accurate representation of students' performance levels, facilitating targeted interventions and informed decision-making to support students' academic growth and achievement.

Measuring Higher-Order Thinking Skills

The study of mathematics is one of the most important disciplines that require more focus and effort in Indonesia. It is cause students in Indonesia need a fundamental understanding of mathematics to have a successful learning experience that will allow them to pursue further education. Students' abilities in critical, logical, and systematic thinking must be developed, and mathematics is one of the tools in scientific thinking required for this development. It is expected of students that they would be able to analyze a problem, come to appropriate conclusions, and think systematically.

Students at all educational levels and in all types of classes, one of which is mathematics class, have a significant demand for the cognitive ability known as HOTS ([Hadi et al., 2018](#)). HOT entails problem-solving, critical, and creative thinking, reasoning, and decision-making. These are a few fundamental skills that today's students are expected to acquire. Thus, it is essential to teach students how to solve HOT situations and refresh their problem-solving thought processes is essential. According to [Brookhart \(2010\)](#) higher-order thinking has been classified into three different contexts of understanding. These contexts include: (1) higher-order thinking as a transfer (students can apply their knowledge and skills which they can further develop in a new context); (2) higher-order thinking as critical thinking (expressing self-reasoning, responding, and decision making without the teacher's intervention); and (3) higher-order thinking as a problem solving (serving students' ability to identify and solve their problems in work and daily).

The determination of dimensions and indicators is based on operational definitions. The operational definition of HOTS is narrowed down to critical thinking skills. The dimension of critical thinking skills is anything that can make someone think critically, namely using concepts, principles, impact prediction, problem-solving, and decision making ([Kania et al., 2022](#)). Therefore, after the researcher has analyzed the different aspects of critical thinking, the indicators of critical thinking are the following: having basic skills, providing explanations, coming to the necessary conclusions, questioning concepts, analyzing concepts, synthesizing relationships between concepts, determining assumption results, using excellent and balanced concepts, locating the source of the problem, making an educated guess as to the cause of the

problem, gathering information to solve the problem, looking for patterns in the data, and looking at alternatives. The dimension of creative thinking ability is everything that can make people think creatively. After analysis by the researcher, the dimensions of creative thinking are working within the limits of competence, trying new things, divergent (spreading) mindsets and imaginative mindsets. Available thinking, not being rigid, developing concepts, modifying concepts, the approach based on trial and error, having original ideas, and having new ideas are all indicators of creative thinking. Rejecting standard techniques, optimizing knowledge, high motivation and content, broad interest, future orientation/optimism, liking challenges/new ideas, and having original ideas are also creative thinking indicators.

In general, the creation of this testing instrument is credible, applicable, and quite doable, and it possesses the potential to have consequences that can be implemented on an ongoing basis. Can be seen from the analysis shown above, most students can improve their numeracy skills by working through the problems provided. It is possible to conclude, after examining the students' responses to the questions presented to them, that the testing instrument can impact the student's ability to solve mathematical problems. However, the learning centered on the student and of making students active participants in their own education is necessary to develop student's knowledge of concepts and numeracy abilities (Nurjanah et al., 2020; Sa'dijah, 2016) This experience informs the design of new approaches to numeracy teaching and assessment (Pettigrew et al., 2020). Although according to Geiger (2018), the subject of research on numeracy encompasses an extensive range of subfields, each of which contains a plethora of significant questions that need to be answered. Instrument development, particularly as it relates to HOTS in mathematics, can be utilized as recommendations and guidelines for the building of learning environments in the classroom.

Higher-order thinking skill

Krulik & Rudnick (1999) describe that HOTS requires understanding, concluding, relating facts and other concepts, in new ways, and applying them find new solutions to problems. Thomas & Thorne (2009) described HOTS is thinking at a higher level that just remembering facts or retelling something (Crumpler, 2014). HOTS go beyond simple recall and reproduction of information. The preceding opinion leads one to conclude that HOTS is a complicated thinking process involving the most fundamental mental functions. This conclusion may be reached since HOTS is used to describe the material, draw conclusions, construct representations, and analyze.

Murray (2013) states that skills that fall under the category of higher-order thinking include critical thinking, the ability to solve problems, creative thinking, and decision making. Besides, Conklin (2012) and Heong (2016) states that higher-order thinking skills include critical thinking and creative thinking. In the revision of Bloom's taxonomy, the cognitive processes that occur when students are active are analyzing, evaluating, and creating (Conklin, 2012). Students' ability to engage in higher-order thinking may improve if they engage in this cognitive process. Higher-order thinking skills, such as critical, logical, reflective, metacognitive, and creative thinking, are all defined there (King et al., 2013). The abilities are active when faced with unfamiliar problems, uncertainties, questions and choices. Higher-order thinking skills are not only applied at school, but also applied in real life (Williams, 2015).

CONCLUSION

The instrument meets the requirements of validity, effectiveness, and practicality. After using it, there is a sharp increase in higher-order thinking skills in critical thinking. The validated and effective nature of the instrument, along with its practicality, establishes a strong foundation for enhancing critical thinking skills, particularly in higher-order thinking. The recommendation to employ similar step-by-step developmental instruments, especially within the context of mathematics learning strategy classes, carries the potential to empower educators in their pursuit of constructing effective assessment tools. By implementing these

developmental steps, educators may gain valuable insights into instrument design, fostering the creation of robust evaluation methods. Furthermore, the authors' belief in the potential contribution of this research to the realm of measuring students' higher-order thinking skills (HOTS) suggests a positive trajectory for future instrument development. As the findings pave the way for improved assessment techniques, the study's implications ripple beyond the immediate context. The exploration of new avenues for research, prompted by this study's outcomes, adds to the growing body of knowledge in the field. Ultimately, this comprehensive exploration of instrument validity and its tangible impact on critical thinking skills lays the groundwork for enhanced educational practices and continued advancements in measurement methodologies.

DECLARATION

Author Contribution : NK: Conceptualization, Writing - Original Draft, Methodology, Editing and Visualization, Writing - Review & Editing, and Formal analysis.
FMK and LMA: Validation and Supervision

Funding Statement : -

Conflict of Interest : The authors declare no conflict of interest.

Additional Information : Additional information is available for this paper.

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>
- Balzano, B. (1991). Curriculum reform activity science and mathematics. *Educational policy and planning project*. 1–26.
- Basuki, I., & Hariyanto. (2014). *Asesemen pembelajaran [Learning assessment]*. Bandung: PT. Remaja Rosdakarya.
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Library of Congress Catalog.
- Brookhart, S. M., & Nitko, A. J. (2011). Strategies for Constructing Assessments of Higher-Order Thinking Skills. *Assessment of Higher Order Thinking Skills*, 1, 327–59.
- Conklin, W. (2012). *Strategies for developing higher-order thinking skills: Grades 6-12*. Shell Educational Publishing, Inc.
- Crumpler, W. (2014). *Higher order thinking skills in critical and creative thinking*. NC State University.
- Desstya, A., Prasetyo, Z. K., Suyanta, Susila, I., & Irwanto. (2019). Developing an instrument to detect science misconception of an elementary school teacher. *International Journal of Instruction*, 12(3), 201–218. <https://doi.org/10.29333/iji.2019.12313a>
- Geiger, V. (2018). Generating ideas for numeracy tasks across the curriculum. *Proceedings of the 41st Annual Conference of the Mathematics Education Research Group of Australasia, July*, 314–321. <https://eric.ed.gov/?id=ED592434>
- Hadi, S., Retnawati, H., Munadi, S., Apino, E., & Wulandari, N. F. (2018). The difficulties of high school students in solving higher-order thinking skills problems. *Problems of Education in the 21st Century*, 76(4), 520–532. <https://doi.org/10.33225/pec/18.76.520>
- Heong, Y. M. (2016). The role of higher order thinking skills in green skill development. *In MATEC Web of Conferences (Vol. 70)*. <https://doi.org/10.1051/mateconf/20167005001>
- Hopson, M. H., Simms, R. L., & Knezek, G. A. (2001). Using a technology-enriched environment to improve higher-order thinking skills. *Journal of Research on Technology in Education*, 34(2), 109–119. <https://doi.org/10.1080/15391523.2001.10782338>
- Ichsan, I. Z., Hasanah, R., Ristanto, R. H., Rusdi, R., Cahapay, M. B., Widiyawati, Y., & Rahman, M. M. (2020). Designing an innovative assessment of HOTS in the science learning for the 21st century. *Jurnal Penelitian dan Pembelajaran IPA*, 6(2), 211. <https://doi.org/10.30870/jppi.v6i2.4765>
- Kania, N., & Juandi, D. (2023). Does self-concept affect mathematics learning achievement?. *Journal of Education and Learning (EduLearn)*, 17(3), 455–461. <https://doi.org/10.11591/edulearn.v17i3.20554>
- Kania, N., Santoso, E., & Nurbelayanti, N. (2022). Bagaimana kemampuan berpikir kritis matematis dan aktivitas siswanya pada model pembelajaran berbasis masalah? [What are the mathematical critical thinking abilities and activities of students in the problem based learning model?]. *Journal on Mathematics Education Research*, 3(2), 27–34.
- King, F., Goodson, L., & Rohani, F. (2013). *Higher order thinking skills: Definition, teaching strategies, assessment*. Tallahassee. Florida State University, Florida.
- Korkmaz, S., Kazgan, A., Çekiç, S., Tartar, A. S., Balcı, H. N., & Atmaca, M. (2020). The anxiety levels, quality of sleep and life and problem-solving skills in healthcare workers employed in COVID-19 services. *Journal of Clinical Neuroscience*, 80, 131–136. <https://doi.org/10.1016%2Fj.jocn.2020.07.073>

- Krulik, S., & Rudnick, J. A. (1999). *Innovative Tasks to improve critical and creative thinking skills*. In I. stiff & f. curcio (Eds.), *Developing Mathematical Reasoning in Grades K-12*. National Council of Teachers of Mathematics.
- Lyons, P., & Bandura, R. P. (2020). Skills needs, integrative pedagogy and case-based instruction. *Journal of Workplace Learning*, 32(7), 473–487. <https://doi.org/10.1108/JWL-12-2019-0140>
- Mardapi, D. (2008). *Technique of preparation of test and non test instrument*. Mitra Cendekia Press.
- Miri, B. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in Science Education*, 37(4), 353–369. <https://doi.org/10.1007/s11165-006-9029-2>
- Moussa, A., Barnett, E. A., Brathwaite, J., Fay, M. P., Kopko, E., Senior, A. M., & Assistant, R. (2020). *A changing paradigm in high school mathematics*. Columbia University.
- Murray, J. (2013). A qualitative synthesis of factors influencing maintenance of lifestyle behaviour change in individuals with high cardiovascular risk. *BMC Cardiovascular Disorders*, 13. <https://doi.org/10.1186/1471-2261-13-48>
- Nurjanah, S., Pebianti, V., & Handaru, A. W. (2020). The influence of transformational leadership, job satisfaction, and organizational commitments on Organizational Citizenship Behavior (OCB) in the inspectorate general of the Ministry of Education and Culture. *Cogent Business & Management*, 7(1), 1793521. <https://doi.org/10.1080/23311975.2020.1793521>
- Organisation for Economic Co-operation and Development. (2015). *PISA 2015: Results in focus*. PISA.
- Ozmantar, M. F. (2017). A historical analysis of primary mathematics curricula in terms of teaching principles. *International Journal of Research in Education and Science*, 3(2), 327–342. <https://doi.org/10.21890/ijres.327890>
- Pettigrew, C., Soldan, A., Wang, J., Wang, M.-C., Arthur, K., Moghekar, A., Gottesman, R. F., & Albert, M. (2020). Association of midlife vascular risk and AD biomarkers with subsequent cognitive decline. *Neurology*, 95(23), e3093–e3103. <https://doi.org/10.1212/WNL.0000000000010946>
- Priyanto, D., & Dharin, A. (2021). Students' creativity development model and its implementation in Indonesian Islamic elementary school. *Pegem Egitim ve Ogretim Dergisi*, 11(3), 81–87. <https://doi.org/10.14527/pegegog.2021.00>
- Retnawati, H. (2014). *Membuktikan validitas instrumen penelitian [Proving the validity of research instruments]*. 31. [http://staffnew.uny.ac.id/upload/132255129/pengabdian/7 Validitas Instrumen-alhamdulillah.pdf](http://staffnew.uny.ac.id/upload/132255129/pengabdian/7%20Validitas%20Instrumen-alhamdulillah.pdf)
- Sa'dijah, C., Rafiah, H., Gipayana, M., Qohar, A., & Anwar, L. (2017). Asesmen pemecahan masalah open-ended untuk mengukur profil berpikir kreatif matematis siswa berdasar gender [Open-ended problem solving assessment to measure students' creative mathematical thinking profiles based on gender]. *Sekolah Dasar: Kajian Teori dan Praktik Pendidikan*, 25(2), 147–159. <http://dx.doi.org/10.17977/um009v25i22016p147>
- Salihu, L., Aro, M., & Räsänen, P. (2018). Children with learning difficulties in mathematics: Relating mathematics skills and reading comprehension. *Issues in Educational Research*, 28(4), 1024–1038. Retrieved from <https://www.iier.org.au/iier28/salihu-abs.html>
- Santoso, S. (2006). *Seri solusi bisnis berbasis TI: Menggunakan SPSS untuk statistik multivariat [IT based business solutions series: using spss for multivariate statistics]*. Elex Media Komputindo.
- Setiawan, J., Sudrajat, A. A., Aman, A., & Kumalasari, D. (2021). Development of higher order thinking skill assessment instruments in learning Indonesian history. In *International Journal of Evaluation and Research in Education (IJERE)* 10(2) p. 545). Institute of Advanced Engineering and Science. <https://doi.org/10.11591/ijere.v10i2.20796>
- Surapranata, S. (2009). *Analisis, validitas, reliabilitas, dan interpretasi hasil tes implementasi kurikulum 2004 [Analysis, validity, reliability and interpretation of the 2004 curriculum implementation test results]*. Bandung: PT. Remaja Rosdakarya.
- Tanudjaya, C. P., & Doorman, M. (2020). Examining higher order thinking in Indonesian lower. *Journal on Mathematics Education*, 11(2), 277–300. <https://doi.org/10.22342/jme.11.2.11000.277-300>
- Tanujaya, B., Mumu, J., & Margono, G. (2017). The relationship between higher order thinking skills and academic performance of student in mathematics instruction. *International Education Studies*, 10(11), 78. <https://doi.org/10.5539/ies.v10n11p78>
- Tasman, F. (2020). Improving junior high school teacher ability in designing higher order thinking (HOTS) problems. In *Pelita Eksakta* 3(1) p. 25). <https://doi.org/10.24036/pelitaeksakta/vol3-iss1/87>
- Taştan, M., & Yılmaz, K. (2008). Organizational citizenship and organizational justice scales' adaptation to Turkish. *Egitim ve Bilim*, 33(150), 87–96. Retrieved from <http://egitimvebilim.ted.org.tr/index.php/EB/article/view/633/>
- Thomas, & Thorne. (2009). *How to increase higher order thinking*. Center for Development and Learning.
- Wang, S., & Wang, H. (2011). Teaching higher order thinking in the introductory MIS course: a model-directed approach. *Journal of Education for Business*, 86(4), 208–213. <https://doi.org/10.1080/08832323.2010.505254>
- Wijaya, A. (2016). Students' information literacy: A perspective from mathematical literacy. *Journal on Mathematics Education*, 7(2), 73–82. <https://doi.org/10.22342/jme.7.2.3532.73-82>
- Williams, R. (2015). *Higher-order thinking skills: challenging all students to achieve*. Skyhorse Publishing.