



Development of Mathematical Reasoning Tests Based on Minimum Competency Assessments with Bengkulu Contexts

Edi Susanto¹, Agus Susanta^{2,*}, & Nur Aliyyah Irsal¹

¹Department of Mathematics Education, University of Bengkulu, Indonesia

²Department of Magister Mathematics Education, University of Bengkulu, Indonesia

Abstract: The achievement of high school graduates' competence is continuously improved by the government in Indonesia, one of which is by implementing the Minimum Competence Assessment. The main problem in its application is the limitations of similar problem models provided in classroom learning so that students are not accustomed to solving them. The purpose of this research is to produce mathematical reasoning questions using the Minimum Competence Assessment model for junior high school students that are valid, reliable, and have a potential impact. This research is development research with two stages of developing Tessmer's model, namely: (1) Preliminary and (2) formative evaluation stage, which consists of self-evaluation, prototyping (expert reviews and small group), and field tests. The trial subjects of this study were 25 students of class VIII Bengkulu state junior high school 7 in Bengkulu City. The results of the study obtained questions that met valid criteria with an Aiken index each of which was more than 0.5. The questions meet reliable criteria with a count of more than 0.6 and are in the high category. The questions have a potential impact with sufficient students' selection abilities on average and students giving a good response of more than 50.00%. The suggestion from the research is that the teacher encourages choosing a real context to present the problem in mathematics learning.

Keywords: development research, instrument test, mathematical reasoning, minimum competence assessment.

Abstrak: Capaian kompetensi lulusan sekolah menengah terus ditingkatkan oleh pemerintah di Indonesia, salah satunya dengan menerapkan Asesmen Kompetensi Minimum (AKM). Permasalahan utama dalam penerapannya adalah keterbatasan model soal serupa yang disediakan dalam pembelajaran di kelas sehingga siswa tidak terbiasa menyelesaikannya. Tujuan dari penelitian ini adalah untuk menghasilkan soal matematika penalaran model AKM untuk siswa SMP yang valid, reliabel, dan memiliki dampak potensial. Penelitian ini merupakan penelitian pengembangan dengan dua tahap pengembangan model Tessmer, yaitu: (1) Preliminary, dan (2) tahap formative evaluation yang terdiri dari: self-evaluation, prototyping (expert reviews dan small group) serta field test. Subjek uji coba penelitian ini adalah siswa kelas VIII SMP Negeri 7 Kota Bengkulu sebanyak 25 orang. Hasil penelitian diperoleh dua paket soal yang memenuhi kriteria valid dengan indeks Aiken masing-masing lebih dari 0,5. Soal memenuhi kriteria reliabel dengan rhitung lebih dari 0,6 dan kategori tinggi. Soal memiliki dampak potensial yang dengan kriteria kemampuan siswa secara rata-rata cukup dan siswa memberikan respon baik lebih dari 50,00%. Saran dari penelitian adalah guru hendaknya memilih konteks nyata untuk menyajikan masalah dalam soal.

Kata kunci: penelitian pengembangan, instrument tes, penalaran matematis, asesmen kompetensi minimum.

▪ INTRODUCTION

The objectives of learning mathematics for junior high schools in Indonesia refer to the Regulation of the Minister of Education and Culture Number 58 of 2014, namely understanding mathematical concepts and the interrelationships between concepts and using reasoning on patterns and characteristics, performing mathematical manipulations in making generalizations, compiling evidence, or explaining mathematical ideas and statements (Mendikbud, 2014). This supports the demands of today's human resource needs, who must master various abilities and skills to survive the development of technology and information. In the 21st century, the skills needed by students are communication skills, collaboration, complex thinking, creativity, innovation, and problem-solving skills (Soulé & Warrick, 2015; Redhana, 2019), HOTS skills (Tan & Haili, 2015).

The achievement of mathematics learning outcomes in Indonesia, especially in secondary schools, still needs to be improved. The performance of Indonesian students in international mathematics surveys has always been in a low category. Referring to the 2019 PISA results, only 0.1% of Indonesian students can solve problems at level 2 (Stacey, 2010). In 2018, Indonesian student scores only reached an average of 379 out of a maximum score of 600, with a rank of 73 out of 79 participating countries (OECD, 2019; Puspendik, 2019). Another survey data, the Trend in International Mathematics and Science Study (TIMSS), shows that the average score of Indonesian students in mathematics is 397, which is lower than the international average score of 500 (Mullis et al., 2012). In Indonesia itself, evaluation on a national basis through computer-based national assessment (in Indonesia ANBK) as a substitute for the national exam shows that results in the field of mathematics are still low. The computer-based national assessment analysis results show that less than 50% of students have reached the minimum competency limit for numeration (Kemdikbud, 2022).

The findings of previous studies show that the ability of high school students in mathematics is still low. Several studies have shown that students' ability to solve questions requires thinking skills in the high category on average in the low category (Susanto & Retnawati, 2016). Other research that has been conducted on middle school students regarding students' ability to complete TIMSS model questions shows that student mastery of reasoning-type questions is 58.33% at a low level and only 8.33% at a high level (Susanta et al., 2021). The results of other research analyses found that at the level of reasoning ability, the student's mastery was the lowest. According to Rosnawati's research (2013), the average lowest percentage achieved by students in the cognitive domain is the reasoning level of 17%. Further research by Susanta, Sumardi, and Susanto (2022) states that the average ability of junior high school students in Bengkulu in solving PISA literacy questions is at level 2.

One of the problems with the low achievement of student mathematics learning outcomes in international surveys and previous research that has been described is that students need to become more familiar with the questions given. This requires the teacher to bridge students so that they are accustomed to solving non-routine questions. This opinion is supported by Leung's statement (Shadiq, 2007), Which states that the emphasis on learning by teachers in Indonesia is more on mastering basic skills, still less emphasis is placed on applying mathematics in the context of everyday life, communicating mathematically, and reasoning mathematically. Research conducted by

Khan (2011) found that teachers rarely give high-level questions. In the application of the 2013 curriculum learning, the questions tested were not able to train students' high-level thinking skills (Fariah et. al, 2018).

Along with the problems related to the achievement of learning outcomes in schools, the government's discourse on implementing minimum competency assessment at every level of education is of particular concern to educators. As educators, they should be able to prepare evaluation tools that are similar to the minimum competency assessment questions provided by the government. Evaluation is one of the most important activities in learning to determine the achievement of learning outcomes that have been carried out (Susanta & Susanto, 2021). However, reality shows that most teachers need to understand this fully. The results of an interview with one of the Bengkulu City Middle School mathematics teachers revealed that the teacher was not used to compiling evaluation tools as a means of training both independently and guided in class related to questions of the minimum competency assessment type.

Questions can be developed based on a minimum competency assessment at the reasoning level for the high school level in compiling the instrument using the context of Bengkulu. The research that has been done proves the role of the regional context in supporting student skills. The use of context in learning to generalize abstract concepts can be understood based on thoughts that are built based on certain realistic situations that are well known to students (Susanti, 2016). Students' thinking skills can be made through practice in solving real problems of everyday life (Warisdiono, 2017). Mathematical problems that use context can present situations that have been experienced in real terms for students (Zulkardi, 2006). The use of regional or cultural contexts can be used as a learning resource. Culture can be in the form of various ideas so that it can be a source of contextual mathematics teaching and learning activities (Sutrimo, Kamid, & Saharudin, 2019).

Several studies on the development of test instruments used as a reference in this study have been carried out. Research by Kamaliyah, Zulkardi, and Darmawijoyo (2013) about developing PISA level 6 questions, Annisah (2011) about PISA levels 2 to 6, developing TIMSS Reasoning Type questions (Susanti, 2016). PISA questions in the context of Indonesia's natural and cultural heritage (Oktiningrum, Zulkardi, & Hartono, 2016). Research on the development of questions using context has also been widely carried out and has become a reference in this development research. The study of Yuliani, Alfarisa, and Tiurlina (2022) developed HOTS questions within the Banten cultural context. Several studies on the development of test instruments used as references in this study have been carried out. Kamaliyah, Zulkardi, and Darmawijoyo (2013) research on the development of PISA level 6 questions, Annisah (2011) on PISA levels 2 to 6, development of TIMSS Reasoning Type questions (Susanti, 2016); Problem type PISA in context of Bangka (Asaprawira, Zulkardi, & Susanti, 2019). Research on the development of questions using context has also been widely carried out and has become a reference in this development research. The study of Yuliani, Alfarisa, and Tiurlina (2022) developed HOTS questions in the context of Banten culture. Problems lime PISA ASEAN GAMES's contexts(Yansen, et.al, 2019; Pratiwi, et.al, 2019).

In contrast to previous studies, the focus on developing questions in this study was mathematics questions based on a minimum competency assessment type of reasoning in the context of Bengkulu.

▪ **METHOD**

Research methods

The method used in this research is the research and development method. The researcher developed a mathematical reasoning problem with a minimum competency assessment type based on the Bengkulu context for junior high school students, which is valid, reliable, and has the potential to impact students' abilities, especially mathematical literacy skills.

Research design

Design development is carried out using the stages of the development model (Tessmer, 1993), which consists of two stages, namely the preliminary stage and the formative evaluation stage. The initial phase consists of the preparation of question designs. The formative evaluation stage includes (a) self-evaluation, (b) prototyping (expert reviews and small groups), and (c) field tests. In the Preliminary Stage, material analysis is carried out based on the 2013 curriculum and an analysis of the Bengkulu context that is by the material resulting from the analysis. In the design stage, indicators are prepared, making grids, writing questions, and preparing scoring guidelines. In the formative evaluation stage, the self-evaluation stage is carried out by assessing and assessing the alignment of the questions with the level of reasoning, the selected material, and the selected Bengkulu context. An expert review carried out an initial prototype assessment focusing on aspects of the material, construction, language, and context used. The assessment was carried out by material experts, namely two lecturers in mathematics at the University of Bengkulu. The results of the revision were based on suggestions from the assessment of the prototype and continued with trials on students. The Small Group stage of the questions that have been validated is tested specifically for the use of questions which include clarity, ease of use, and presentation of the questions. The trial was carried out in small groups of 9 people with different levels of ability (low, medium, high). The field test stage was tested on a wider scale to analyze students' ability to complete AKM questions and measure the reliability of the instrument being developed.

Research subject

This research was conducted on class VIII students of SMP N 11 Bengkulu City in the 2022/2023 academic year. Our limited or small group test subjects chose 9 students in class VIII C with different levels of ability distribution, namely low, medium, and high students. In the broad potential impact test, we chose class VIII A students with a total of 25 students with 16 female and 9 male students.

Research instrument

In collecting data, we do it using observation, tests, and questionnaires. The instruments used were validity sheets and mathematical reasoning test questions. The validity sheet in the form of a statement measures aspects of content, construct, language, and use of context in as many as 12 items with a score of 1-5 on a Linkert

scale. The assessment criteria in the validation sheet consist of: very suitable (5), appropriate (4), sufficient (3), not applicable (2), and wildly inappropriate (1). The student response questionnaire is a statement to measure the student's potential impact after completing the test.

Furthermore, the test instrument is in the form of mathematical reasoning questions that measure students' ability to solve minimum competency assessment questions, which are the product of development in this study. The test instrument is in the form of multiple-choice questions and descriptions with a total of 8 items. The instrument was developed based on the material aspects of minimum competency assessment with material: geometry, numbers, algebra, data, and opportunities (Kemendikbud, 2021). Indicators for the development of reasoning questions refer to NCTM (2000), including recognizing understanding as a fundamental aspect of mathematics; making and deriving mathematical conjectures, developing and evaluating mathematical arguments, and selecting and using different types of reasoning.

Data analysis technique

In this study, data analysis was carried out in a qualitative descriptive manner which explained the quality of the questions we had developed. Data analysis consisted of data validity analysis, question reliability analysis, and analysis of students' ability level in solving questions based on minimum competency assessment. Validity analysis based on experts uses the Aiken validity index (1980) quoted from Retnawati, 2014 with valid criteria if the Aiken value is more than 0.5.

Reliability analysis was performed on multiple choice questions and descriptions separately. In objective questions, we use Kuder and Richardson's analysis, while the alpha coefficient for essay questions (Reynold et al, 2011) with the criteria of questions is said to be reliable if the calculated results are more than 0.6 (Basuki and Hariyanto, 2014). The additional analysis that we use is the standard error measurement (SEM) analysis (Reynold, et al, 2011). Furthermore, data on student test results are described qualitatively to provide information about how students complete and identify student errors given. The criteria for each level of student thinking are based on a conversion value of 0-100 with criteria: 76-100 (very good), 51-75 (good), 26-50 (enough), and a range of 1-25 (poor). The last analysis of the data was carried out by analyzing student responses. Analysis of student response data aims to measure the potential impact of the instrument as measured by student responses after completing the questions. The questionnaire consists of four statements with yes or no options. Data analysis calculated the proportions at each port from wide-scale trials

▪ RESULT AND DISSCUSSION

Description of Product Development Results

This study resulted in a mathematical reasoning test question based on the Minimum Competency Assessment (In Indonesia AKM) initiated by the Indonesian Ministry of Education and Culture in the Bengkulu context. The test questions developed were 8 items with four multiple-choice questions and four description questions. The questions were developed based on the distribution of AKM material (Kemendikbud, 2021): geometry, algebra, numbers, data, and uncertainty, with two questions each. The product questions were developed to refer to the selected Bengkulu context, namely: the context of tourism, the context of historical buildings, culture, and

context of typical food. This study presents results based on the stages of Tessmer's (1993) development model.

Preliminary Stage Result

At the preparation stage, the researcher examined several aspects supporting the development of test questions, namely the Bengkulu material and context. The material was analyzed based on the essential competencies in the 2013 curriculum by grouping the material: numbers, algebra, geometry, and data and opportunities. Each material is reviewed for basic competencies that allow it to be designed according to the characteristics of the reasoning questions that refer to the reasoning questions on the TIMSS. At the initial stage, the completeness of the development of questions is also designed, such as lattice questions and reference questions for assessment.

*Formative Evaluation Stage
Self-evaluation*

At this stage, the author assesses the suitability of the context with the selected material. This stage also analyzed the Bengkulu context by the material being developed. The results of the context analysis obtained four basic contexts that were by the development material, namely geometry material with the context of historic buildings and typical food, number material with the context of tourism and typical food, algebraic material with the context of culture and tourism objects, and data and opportunities material with object contexts, tour. The result of this stage is the arrangement of question indicators, problem grids, and initial draft questions. The following is a summary of the initial draft of the development product.

Table 1. Overview of the initial draft

Type	Question	Material	Bengkulu Context
Multiple choice	1	Numbers	Bengkulu Toursm
	2	Algebra	Bengkulu Culture
	3	Geometry	Special Food
	4	Data and Chance	Bengkulu Tour
Essay	1	Numbers	Special Food
	2	Aljgebra	Tourist Object
	3	Geometry	Historical Building
	4	Data and Chance	Bengkulu Culture

At this stage, checking the draft questions that have been prepared related to conformity with the level of reasoning.

Expert Review

This draft was assessed by two validators who are lecturers in the mathematics education master's program-the assessment results of two validators on material, construction, language, and context use. Assessment data from the validator with a rating scale of 1-5 were analyzed to calculate the Aiken index, which is summarized in the following table.

Table 2. Validity result

Question Type	Item	Mean Aiken' V	Criteria
Multiple choice	1	0.64	Valid
	2	0.72	Valid
	3	0.68	Valid
	4	0.60	Valid
Essay	1	0.65	Valid
	2	0.56	Valid
	3	0.68	Valid
	4	0.59	Valid

Based on the analysis of the Aiken index for each question that has been developed, on average, the questions fulfill the validity aspect. This is because the valid criteria are fulfilled where each Aiken' V has a value of more than 0.5. This means that the questions developed are theoretically appropriate and refer to construction, content, and language aspects. The importance of validity aspect of a learning product or learning measurement tool so that the information obtained from the instrument is appropriate. This is the opinion of Rezeki et al. (2022), who states that valid learning tools are the best way to obtain information about student abilities by learning objectives. The development of educational products must pay attention to the validation process because it is very important to get the best quality educational products (Risnawati et al., 2019).


The validator obtains suggestions and inputs at the expert validity stage. The following summarizes the validator's recommendations for the designed questions. Based on the suggestions the validator gave, revisions were made to the product questions being developed.

Table 3. Summary of validator suggestions

Expert Rivew	Revise
Validator 1	<ul style="list-style-type: none"> ▪ In using context, it is better to accompany pictures or supporting photos so that students easily understand it ▪ Improve the editorial sentence statement in point a so that students can complete it
Validator 2	<ul style="list-style-type: none"> ▪ Add illustrations to pictures ▪ Correct the sentence redaction on the question ▪ The language of opportunity should be changed

The following is an example of the results of revising the description questions with data content and opportunities and Tabot Bengkulu contests.

Table 4. Summary of validator suggestions

Before Revision	After Revision
<p>Problem: Tabot [Data and Chance] At the 2021 Tabot event in the city of Bengkulu, there is a Ferris wheel that has 8 cages, each cage containing two passengers. Each cage has certain colors as follows: Yellow = 1 cage Red = 3 cages Green = 2 cages Blue = 2 cages.</p> <p>Questions</p> <ol style="list-style-type: none"> a. What color cage does the passenger get off first? b. Order the probability of a cage whose occupants descend from largest to smallest! 	<p>Problem: Tabot [Data and Chance] At the Tabot event in 2021 in the city of Bengkulu, there is a Ferris wheel game that has 8 cages, and each cage can be occupied by two passengers, as shown in the following illustration.</p> <div style="text-align: center;">  </div> <p>Figure Tabot and Ferris wheel (Source: journalelite.co.id)</p> <p>Each cage has certain colors as follows: Yellow = 1 cage Red = 3 cages Green = 2 cages Blue = 2 cages.</p> <p>Question: After the Ferris wheel rotates, determine:</p> <ol style="list-style-type: none"> a. What color cage has the first occupant alighted based on the greatest probability? a.b. The order of the possible cages whose occupants descend from largest to smallest!

Small Group Stage

The small group stage involves 9 students with characteristics of the high, medium, and low ability levels, each consisting of three students. In carrying out trials in small groups, this was carried out by working on questions by the subject. After working on the questions, students were asked to assess by giving responses to the legibility of the questions assessment of students through a readability questionnaire. The results of this stage are in the form of student responses, namely: students provide answers to questions that are easy to understand, the context used in the questions is known by students, the presentation of symbols is easy to read, and the writing used is easy to read.

Field Test Stage Results

Test product questions that were developed extensively in large-scale classes with 25 students. We use the data from the large-scale trial to test the questions' reliability and the standard error measurement (SEM) to support the quality of the developed

instrument. The results of calculating the reliability of MCA-based reasoning math questions on objective questions with a reliability value (R11) of 0.802 (high category). The data showed that the standard deviation in multiple-choice questions was 0.504, so the SEM was 0.224. In the description questions, the Cronbach alpha value is 0.628 (high category) with a standard deviation of 0.94, with an SEM of 0.573.

Potential Effect

Data recap of students' ability in solving AKM questions descriptively is summarized as follows.

Table 5. Category of student ability

Category	Students	Percentage
Excelent	1	4.00%
Good	3	12.00%
Moderate	12	48.00%
Low	9	36.00%

Based on Table 5 it can be seen that students' ability to solve mathematical reasoning questions of the AKM type is in the sufficient category with a percentage of 48.00%. While only one student scored on the excellent criteria with a percentage of 4.00%. In the low category as many as 9 students or 36.00%. The following is an example of a student's answer to one of the developed minimum competency assessment-type reasoning math questions.

Analysis of student answers, in general, showed that students' mastery of questions was still in the low to sufficient category. This shows that there is still a need to emphasize similar questions in learning. We present one example of a student's answer in solving the problem in the developed question [type of question: description, material: data and uncertainty, context: Bengkulu culture].

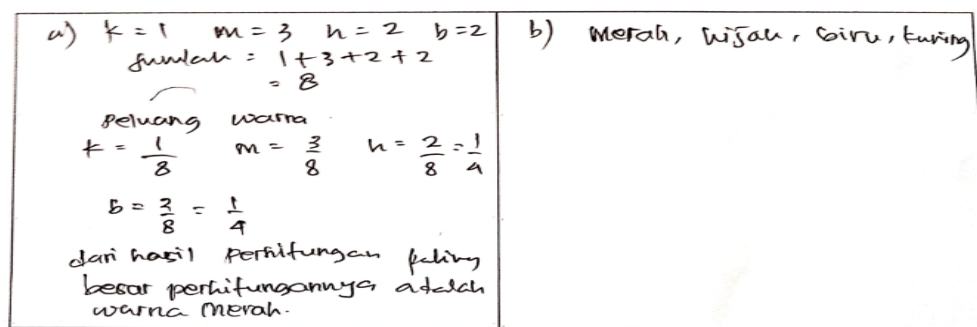


Figure 1. Examples of student answers

Based on Figure 1, the students' answers have been answered correctly, but in point part b the stages of completion have not been explained. Based on the description of the results of student answers and analysis of student answer sheets, it shows that students in general have moderate abilities. These results indicate that some students have been able to answer the questions correctly. The potential impact of this question is also measured through student response questionnaires. The results of student

responses after working on PISA-type reasoning math problems with the Bengkulu context are shown in the following table.

Table 6. Student response to questions

Interest in questions	Respon of students (%)
I am very interested in the problem in the problem	64.00
seriously solved the problem with the problem we are currently in	60.00
I'm not interested in solving problems	12.00
Context makes it easier for me to solve problems	52.00

Based on Table 6, it is known that the average student gave a good response to the questions developed. Thus, it can be concluded that the developed test instrument has a potential impact on students' reasoning abilities. The use of real contexts or local contexts that are close to students helps students in modeling problems in mathematics. The results of this study are supported by several previous studies. Zulkardi and Ilma (2006) suggest that mathematical problems that use various contexts will be able to present real situations that have been experienced for children. Javanese traditional games are potential to be an alternative media in improving children's problem-solving skills (Iswinarti & Suminar, 2019). Problems used in learning must be close to students (Susanta, Sumardi, & Zulkardi, 2022). PISA model math problems content change and relationships have a potential effect on junior high school students (Ahyan, Zulkardi, & Darmawijoyo, 2014).

▪ **CONCLUSION**

This study resulted in a mathematical immersion of PISA-type reasoning that was valid, reliable, and had a potential impact on students. The results of the validity analysis with the Aiken index show that each question meets the valid criteria with an Aiken V of more than 0.5. The results of the reliability test on the purpose of the question obtained an R11 value of 0.802 (high category). From the data, the standard deviation of multiple-choice questions is 0.504 so the SEM is 0.224. In the description problem, the Cronbach alpha value is 0.628 (high category) with a standard deviation of 0.94 and the SEM is 0.573. Questions have a potential impact on students' reasoning abilities.

Based on the results of the development of the questions that have been carried out, suggestions that can be given by the author are that teachers need to use context in preparing instrument tests. In using the context, the teacher mostly chooses a context that is appropriate to the material and is clearly described so that students can easily understand the problem in the problem.

▪ **REFERENCES**

Ahyan, S., Zulkardi, Z., & Darmawijoyo, D. (2014). Developing mathematics problems based on PISA level of change and relationships content. *Journal on Mathematics Education*, 5(1), 47-56.

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and psychological measurement*, 40(4), 955-959.

- Annisah. (2011). Pengembangan soal matematika model pisa pada konten quantity untuk mengukur kemampuan penalaran matematis siswa sekolah menengah pertama [development of pisa model mathematical problems on quantity content to measure student's mathematical reasoning ability of junior high school]. In *Prosiding Seminar Nasional Pendidikan 2011*. Fakultas Keguruan dan Ilmu Pendidikan, Universitas Sriwijaya. Palembang, 7 Mei 2011
- Basuki, I., & Hariyanto. (2014). *Asesemen pembelajaran [Learning assessment]*. Bandung: PT. Remaja Rosdakarya.
- Dasaprawira, M. N., & Susanti, E. (2019). Developing Mathematics Questions of PISA Type Using Bangka Context. *Journal on Mathematics Education, 10(2)*, 303-314.
- Farihah, N., Imanah, U. N., & Hidayati, E. W. (2018). Pengembangan soal higher order thinking skills (hots) pada materi barisan dan deret bilangan [development of higher order thinking skills (hots) questions on sequences and sequences of numbers]. *MAJAMATH: Jurnal Matematika dan Pendidikan Matematika, 1(2)*, 142-154.
- Iswinarti, I., & Suminar, D. R. (2019). Improving children's problem-solving skills through javanese traditional games. *Cakrawala Pendidikan, 38(3)*.
- Kamaliyah, K., Zulkardi, Z., & Darmawijoyo, D. (2013). Developing the sixth level of PISA-like mathematics problems for secondary school students. *Journal on Mathematics Education, 4(1)*, 9-28.
- Kemendikbud. (2014). Peraturan menteri pendidikan dan kebudayaan, nomor 58, tahun 2014, tentang kurikulum 2013 sekolah menengah pertama (smp)/madrasah tsanawiyah (mts) [regulation of the minister of education and culture, number 58, year 2014, regarding the 2013 curriculum for junior high schools]
- Kemendikbud. (2021). *Framework asesmen kompetensi minimum (akm) [minimum competency assessment framework]*. Jakarta: Pusat Asesmen dan Pembelajaran, Badan Penelitian, Pengembangan dan Perbukuan, Kementerian Pendidikan dan Kebudayaan.
- Kemendikbud. (2022). *Rapor pendidikan publik 2022 [public education report 2022]*. Ppusmendik.Kemdikbud
- Khan. W.B. (2011). *A study of lower-order and higher-order questions at secondary level*. Pakistan: Canadian Center of Science and Education. Asian Social Science, 7(9). September 2011.
- Martin, M. O., Mullis, I. V., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 international results in science*. international association for the evaluation of educational achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.
- Nitko, A.J. & Brookhart, S.M. (2011). *Educational assessment of students*. Xth edition. Upper Saddle River, New Jersey: Prentice Hall Englewood Cliffs.
- NCTM. (2000). *Principles and standards for school mathematics*. reston, va:national council of teachers of mathematics.
- OECD. (2019). *PISA 2018. PISA 2018 Result Combined Executive Summaries*. PISA-OECD Publishing
- Oktiningrum, W., Zulkardi, Z., & Hartono, Y. (2016). Developing PISA-like mathematics task with Indonesia natural and cultural heritage as context to assess students mathematical literacy. *Journal on Mathematics Education, 7(1)*, 1-8.

- Pratiwi, I., & Putri, R. I. I. (2019). Long jump in asian games: context of pisa-like mathematics problems. *Journal on Mathematics Education*, 10(1), 81-92.
- Puspendik. 2019. Pusat asesmen dan pembelajaran [Assessment and learning center. Badan Penelitian dan Pengembangan dan Perbukuan Kemendikbud RI Tentang PISA
- Redhana, I. W. (2019). Mengembangkan keterampilan abad ke-21 dalam pembelajaran kimia [Develop 21st century skills in chemistry learning]. *Jurnal Inovasi Pendidikan Kimia*, 13(1).
- Retnawati, H. (2014). Membuktikan validitas instrumen dalam pengukuran [Proving the validity of the instrument in measurement]. Diambil dari <http://www.evaluation-edu.com>.
- Reynolds, C.R, Livingston, R.B., & Willson, V. (2010). Measurement and assessment in education. Second edition. Upper Saddle River, New Jersey: Pearson Education.
- Rezeki, S., Andrian, D., & Safitri, Y. (2021). Mathematics and cultures: A new concept in maintaining cultures through the development of learning devices. *International Journal of Instruction*, 14(3), 375–392.
- Risnawati, Andrian, D., Azmi, M. P., Amir, Z., & Nurdin, E. (2019). Development of a definition mapsbased plane geometry module to improve the student teachers' mathematical reasoning ability. *International Journal of Instruction*, 12(3), 541–560
- Rosnawati, R. (2013). Kemampuan penalaran matematika siswa SMP Indonesia pada TIMSS 2011 [The mathematical reasoning abilities of Indonesian junior high school students in TIMSS 2011]. In *Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA, Fakultas MIPA, Universitas Negeri Yogyakarta* (Vol. 18, pp. 1-6).
- Shadiq, F. (2007). Penalaran atau reasoning. perlu dipelajari para siswa di sekolah [reasoning or reasoning. students need to learn in school]. *Jurnal Online*. https://fadjarp3g.files.wordpress.com/2007/09/ok-penalaran_gerbang_.pdf.
- Soulé, H., & Warrick, T. (2015). Defining 21st-century readiness for all students: What we know and how to get there. *Psikologi Estetika, Kreativitas, dan Seni*, 9(2), 178-186.
- Stacey, K. (2010). The pisa view of mathematical literacy in indonesia. *Journal on Mathematics Education (IndoMS-JME)*. July, 2011, Volume 2.
- Susanta, A., & Susanto, E. (2021). The development of test instruments to support students' ability on mathematical learning evaluation course. In *Journal of Physics: Conference Series* (Vol. 1731, No. 1, p. 012057). IOP Publishing.
- Susanta, A., Sumardi, H., & Susanto, E. (2022). Students' mathematical literacy ability in solving problems on pattern material using the bengkulu local contexts. *Jurnal Didaktik Matematika*, 9(2), 343-357.
- Susanta, A., Susanto, E., Maizora, S., & Rusdi, R. (2021). Analisis kemampuan siswa smp/mts kota bengkulu dalam menyelesaikan soal matematika timss. [analysis of the ability of bengkulu city junior high school/mts students in solving the timss mathematics problems]. *Jurnal THEOREMS (The Original Research of Mathematics)*, 5(2), 131. <https://doi.org/10.31949/th.v5i2.2567>

- Susanti, E. (2016). Pengembangan soal matematika tipe timss menggunakan konteks rumah adat untuk siswa sekolah menengah pertama [Timss type math problem development using traditional house context for junior high school students]. *Jurnal Pendidikan Matematika*, 10(2), 53-74.
- Susanto, E., & Retnawati, H. (2016). Perangkat pembelajaran matematika bercirikan PBL untuk mengembangkan HOTS siswa SMA [Mathematics learning tools are characterized by PBL to develop HOTS for high school students]. *Jurnal Riset Pendidikan Matematika*, 3(2), 189-197.
- Sutrimo, S., Kamid, K., & Saharudin, S. (2019). LKPD bermuatan inquiry dan budaya jambi: efektivitas dalam meningkatkan kemampuan berpikir kreatif matematis. *IndoMath: Indonesia Mathematics Education*, 2(1)
- Warisdiono, et al. (2017). Modul penyusunan higher order thinking skill (HOTS) [Module for compiling higher order thinking skills (HOTS)]. *Jakarta: Direktorat Pembinaan SMA, Direktorat Jenderal Pendidikan Dasar dan Menengah Departemen Pendidikan dan Kebudayaan*.
- Yansen, D., Putri, R. I. I., & Fatimah, S. (2019). Developing pisa-like mathematics problems on uncertainty and data using asian games football context. *Journal on Mathematics Education*, 10(1), 37-46.
- Yen, T. S., & Halili, S. H. (2015). Effective teaching of higher order thinking (HOT) in education. *The Online Journal of Distance Education and e-Learning*, 3(2), 41-47.
- Yuliani, A., Alfarisa, F., & Tiurlina, T. (2022). The development of mathematical HOTS questions based on banten culture. *International Journal of Trends in Mathematics Education Research*, 5(1), 44-52.
- Zulkardi dan Ilma, R. (2006). Mendesain sendiri soal kontekstual matematika [design your own contextual math problem]. *In Prosiding Konferensi Nasional Matematika XIII, 2006*