

Volume 6 Nomor 2, Agustus 2021, halaman 154 – 174.

Development Of Test Instruments To Measure High Order Thinking Skill (Hots) Mathematics Of Students In MTs 2 States Of Medan City

Pengembangan Instrumen Tes Untuk Mengukur *High Order Thinking Skill* (HOTS) Matematis Siswa Mtsn 2 Kota Medan

Qania Agustika Siagian¹, Prihatin Ningsih Sagala²

¹Department of Mathematics Education, Universitas Negeri Medan, qaniaagustika@gmail.com

²Department of Mathematics Education, Universitas Negeri Medan, alfathmommy81@gmail.com

ABSTRAK

Upaya yang dapat dilakukan agar siswa Indonesia mampu ikut bersaing dengan negara-negara lain dalam penilaian TIMSS dan PISA, seharusnya siswa Indonesia harus dibiasakan bertemu dengan instrumen tes yang menuntut High Order Thinking Skill (HOTS). Tujuan pada penelitian ini adalah untuk mengetahui kemampuan berpikir tingkat tinggi matematis siswa, serta mengetahui validitas, reliabilitas, tingkat kesukaran dan daya beda produk instrumen tes berorientasi *High Order Thinking Skill* (HOTS). Penelitian pengembangan ini menggunakan tipe ADDIE. Ada 57 siswa yang dijadikan subjek pada penelitian ini. Hasil penelitian menunjukkan bahwa 15 soal yang telah dianggap berhasil valid dikembangkan, hal ini pada skor rata-rata penilaian instrumen tes oleh ahli sebesar 91% dan validitas setiap butir tes dinyatakan "valid" karena $t_{hitung} \geq t_{tabel}$. Reliabilitas instrumen tes setiap paket soal dinilai reliabel dalam kategori tinggi. Dalam analisis tingkat kesulitan, terlihat bahwa menunjukkan hasil pada kategori mudah dan sedang. Dalam analisis daya pembeda, terlihat bahwa menunjukkan hasil cukup baik dan baik. Hasil rata-rata kemampuan berpikir tingkat tinggi matematis siswa kategori cukup.

Kata Kunci : Instrumen Tes, *High Order Thinking Skill*, ADDIE..

ABSTRACT

Efforts can be made so that Indonesian students can compete with other countries in the TIMSS and PISA assessments, Indonesian students should be accustomed to meeting test instruments that demand High Order Thinking Skills (HOTS). The purpose of this study was to determine the students' high-order mathematical thinking skills, as well as to determine the validity, reliability, difficulty level, and differentiating power of the High Order Thinking Skill (HOTS) oriented test instrument. This developmental research uses the ADDIE type.. There were 57 students who were used as subjects in this study. The results showed that 15 questions that were considered successful and valid were developed, this is the average score of the test instrument assessment by experts of 91% and the validity of each test item is declared "valid" because $t_{count} \geq t_{table}$. The reliability of the test instrument for each package of questions was rated as reliable in the high category. In the analysis of the level of difficulty, it appears that the results are in the easy and medium categories. In the analysis of discriminatory power, it appears that the results are quite good and good. The average results of students' high-order mathematical thinking skills are insufficient category.

Keyword(s): Test Instruments, High Order Thinking Skill, ADDIE..

How to Cite: Siagian, Q. A., & Sagala, P.N. (2021). Development Of Test Instruments To Measure High Order Thinking Skill (Hots) Mathematics Of Students In MTs 2 States Of Medan City. *Mathline: Jurnal Matematika dan Pendidikan Matematika*, Vol. 6 No. 2, 154 - 174.

DOI: <https://doi.org/10.31943/mathline.v6i2.222>

PRELIMINARY

The results of Trends In Mathematics and Science Study (TIMSS) 2015 in Khodaria et al (2019, p.66) Indonesia was ranked 44th out of 49 countries in the world with a math score of 397 with an average of 505. Likewise, the results of the 2018 Program for International Student Assessment (PISA) study showed that Indonesian students were ranked 73 out of 78 countries in the world with a math score of 379 with an average of 489 points (OECD, 2019). Even though the demands of competition in the field of education are very much needed by the younger generation in this era of globalization. Indonesian education must be able to prepare graduates who are able to compete in the world.

Furthermore, other evidence that the higher-order mathematical thinking ability of Indonesian students is still in the low category as the results of Hadi's research (2018) shows that there are 8.33% of students find it difficult to translate problems, 15.59% of students find it difficult to transform problems, 32.53% of students difficulty in the process of working on questions, and 1.34% of students have difficulty in concluding. The results of Ardhana's research (2017, p.16) concluded that the percentage in the category of high order thinking skills which includes analyzing (C4), evaluating (C5), and creating (C6) only reached 2.941%, which is still low.

To reach students from Indonesia who are able to compete with students from other countries, it is necessary to improve the quality of education, which can be built by improving the quality of learning and the quality of the assessment system (Malik, 2018, p.14). The quality of this good learning can be seen from the results of the assessment. Therefore, teachers need to plan lessons, strategies, and question-based constructs that help students think at a higher level, think creatively and think critically. In connection with the structure of the question, the assessment of learning outcomes must be carried out properly, starting from the determination of the instrument, preparation of the instrument, review of the instrument, implementation of the assessment, analysis of the assessment results, and the follow-up program of the assessment results. Efforts can be made so that Indonesian students can compete with other countries in the TIMSS and PISA tests, Indonesian students must get used to meeting questions that are in accordance with the TIMSS and PISA assessment standards. Referring to the 21st-century learning framework, HOTS is the answer in responding to the challenges of 21st-century learning (Rapih & Sutaryadi, 2018, p.77).

The results of observations made by researchers on students at MTs Negeri 2 Medan make the evidence stronger that students' higher-order mathematical thinking skills

are still low. The questions made by the teacher are known to not require all higher-order mathematical thinking skills. Questions made by the teacher can only demand to analyze information from what is known and what is asked from the question.

Anderson & Krathwohl (2001) dalam Motlhabane (2017, p.922) classify cognitive processes into six categories, namely remembering, understanding, applying, analyzing, evaluating, and creating. This is in line with the opinion of Arlianty et al (2018, p.1) based on Bloom's Taxonomy, the categories of remembering, understanding, applying are Lower Order Thinking Skills (LOTS) while analyzing, evaluating, and creating are High Order Thinking Skills (HOTS).

According to Heong et al (2011) in Kusuma (2017, p.26) this higher order thinking ability requires someone to apply information to reach possible answers in new situations. The HOTS assessment instrument as a learning assessment is effective for training students' HOTS and effective in measuring students' thinking skills based on each student's HOTS level. A good test must meet the following criteria well.

1. Validity

Validity is a term related to the extent to which a test has measured what it is supposed to measure. Validity is the ability of a measuring instrument to accurately measure the condition being measured. As Nieveen & Piomp (2007) in Hajar et al (2021, p.4) stated in *An Introduction to Educational Design Research*, the component of material should be based on the state-of-the-art knowledge (content validity) and all components should be consistently linked to each other (construct validity). In this research, the product produced is a high-level mathematical thinking test instrument. To determine the quality of the developed product, it must have validity criteria. To be able to achieve the validity of the test instrument through an expert validation process. The indicator components of the expert validation aspect in general are based on language, construct, and content.

2. Reliability

The concept of reliability or often called reliability (reliable or trustworthy) refers to consistency or homogeneity. This consistency and homogeneity can be seen from the score of the measurement results, both once and several times (Febri E., 2017, p.139).

3. Level of Item Difficulty

The level of difficulty of the item means the degree of a question. Good quality questions are not too difficult and not too easy in other words the questions are proportional (Arifin, 2017, p.266).

4. Differentiating Power of Questions

The calculation of discriminating power is a measurement of the extent to which an item can distinguish between students who have mastered competence and students who have not/less mastered competence based on certain criteria (Arifin, 2017, p.273).

According to Krathwohl (2001) indicators for measuring higher-order thinking include analyzing, evaluating, and creating (Febiana 2019, p.10), which are described below:

(1) Analyzing

- 1a. Analyze incoming information and break down or structure the information into more specific parts to find out patterns/formulas or relationships.
- 1b. Able to recognize and distinguish causal factors and distinguish between causal factors and influencing factors from complex schemes.
- 1c. Identify/formulate the question.

(2) Evaluate

- 2a. Assess solutions, ideas, and methodologies using appropriate criteria or existing standards to determine their effectiveness or usefulness.
- 2b. Generate hypotheses, critique, and test them.
- 2c. Accept or reject a question based on predetermined criteria.

(3) Creating

- 3a. Generalize an idea or perspective about something.
- 3b. Designing a solution to solve a problem.
- 3c. Organize elements or parts into a new structure that has never existed before.

From the existing problems and the provisions for developing a test instrument that can measure students' High Order Thinking Skills (HOTS), this study aims to:

1. Knowing the high-order mathematical thinking skills of class VIII MTs Negeri 2 Medan on the subject matter of Number Patterns, Sequences, and Series.
2. Knowing the validity, reliability, level of difficulty, and differentiating power of the test instrument product-oriented to High Order Thinking Skill (HOTS) on the material of Number Patterns, Sequences, and Series in class VIII SMP/MTs.

METHOD

This research was conducted at MTs 2 States of Medan in the first semester of the 2020/2021 Academic Year with the subjects in this study were 10 students of class VIII-3 as a small class test (development), then for the implementation class trial as many as 29

students of class VIII Plus 1 and 28 students of class VIII Plus 2. The test instrument developed was divided into three packages, each of which amounted to 5 items, namely Question A, Question B, and Question C which were distributed to each student according to their serial number. The reason for dividing the questions into three packages is because it is not efficient for a student to work on 15 HOTS questions within 2 hours of lessons.

In this study, the ADDIE modeling procedure was applied which consisted of five stages, namely (A) analysis, (D) design, (D) development, (I) implementation, and (E) evaluation. The five procedures in the ADDIE model need to be carried out systematically, which are briefly described in stages according to Branch (2009) in Wulan S, et al (2017, p.40) as shown in Figure 1. below:

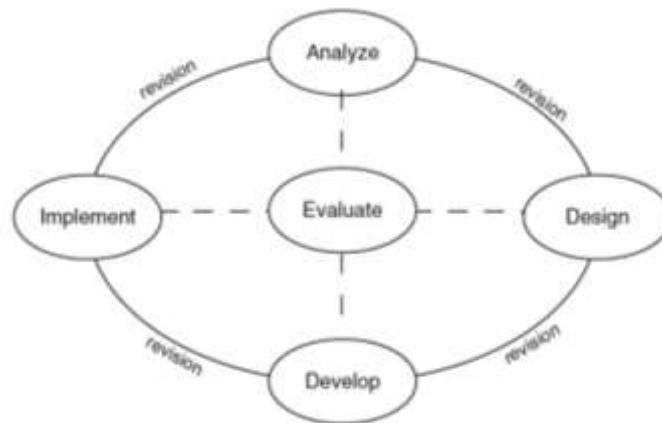


Figure 1. Stages of the ADDIE Model

Based on Figure 1, the first stage is analysis, the researcher analyzes the need for product development, analyzes the feasibility and development requirements, the second stage is planning, the researcher plans the preparation of the test instrument grid, test instrument, and alternative completion of the mathematical High Order Thinking Skill (HOTS) test instrument. In the third stage, namely the development stage, where at this stage the researcher conducts expert validation and analyzes the validation results of the test instruments that have been developed and provides a questionnaire response to the test instrument to the teacher and small class trials. Furthermore, in the fourth stage, namely implementation, the researchers conducted the testing phase of the test instrument in large classes. In the last stage, namely evaluation, researchers evaluate each stage starting from analysis, design, and development. Evaluation can be classified into the formative evaluation and summative evaluation. Evaluation is used to assess the process and test instruments developed.

Research Instruments

To measure the validity, reliability, level of difficulty, and differentiating power of the development of the High Order Thinking Skill (HOTS) test instrument, several research instruments are needed. The instruments used in this study are as follows.

1. The test instrument developed is based on HOTS. The tests that will be given are High Order Thinking Skill questions because to measure students' high-level mathematical thinking skills, questions that include indicators of analyzing, evaluating, and creating are used. The test instrument was used to obtain data about students' higher-order mathematical thinking skills.
2. Expert validation sheet. Some of the validation sheets used are the width of the mathematical High Order Thinking Skill (HOTS) test instrument validation. The results of this test instrument validation sheet are used as a basis for evaluating the developed test instrument.
3. Instrument response of students and teachers. This response sheet is used to obtain data about the statements and opinions of students and teachers on the developed test instrument. This instrument is used when in the development stage of ADDIE, which will be used as the basis for product improvements that will be developed at a later stage.

Test Item Quality Analysis

(1) Expert validation data analysis

In the data from the test instrument validation results, the rating scale criteria will be found, namely: 1 (poor), 2 (good enough), 3 (good), and 4 (very good). Based on the opinion of 5 experts, it can be determined the average value for each aspect, so that the average value of the total aspects is obtained. The activity of determining the average value of the total aspect of assessing the validity of the mathematical High Order Thinking Skill (HOTS) test instrument follows the following steps:

$$\text{score percentage} = \frac{\text{number of scores per indicator}}{\text{maximum number of indicators}} \times 100\% \quad (\text{Fadlila, 2019, p.100})$$

(2) Analysis of the validity of each test item

According to Sugiyono (2017, p.200) to find the coefficient value can be obtained from the product moment correlation formula. The test instrument is said to meet the valid indicators if the higher-order mathematical thinking test instrument developed is at least in the category $85\% < x \leq 100\%$ (can be used without revision).

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

Information:

x : student score on an item

y : student scores on all items

Furthermore, to test the significance of the validity correlation coefficient, it was carried out by means of the t-test, according to Riduwan (2012) in Jannah et al (2017, p.80) with the formula:

$$t_{count} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Information:

r : correlation coefficient, the result of r count

n : many respondents of the trial

The t value obtained from this calculation is then compared with the t value from the table at a certain level of confidence. Distribution (Table t) for $\alpha = 0,05$ and degrees of freedom ($dk = n - 2$). This validity test is carried out on each item with a decision rule if $t_{count} \geq t_{table}$ means valid.

(3) Reliability analysis

The test instrument is said to meet the reliability indicator if the higher-order mathematical thinking test instrument developed is at least in the category $0,6 \leq r \leq 0,8$ (high test reliability). If a test score is reliable, then the score can be used for various purposes (Dewi, 2017). According to Sugiyono (2017, p.27) for the measuring instrument in the form of a description, the Cronbach Alpha formula is used as follows.

$$r = \left[\frac{k}{k-1} \right] \left[\frac{s_t^2 - \sum s_i^2}{s_t^2} \right]$$

Information:

r : reliability coefficient

k : the number of questions

s_i : standard deviation of the i-th test item

s_t : standard deviation of all test items

(4) Difficulty level analysis

The difficulty index (IK) of a test item describes the degree of the proportion of the

number of correct answer scores on the test item in question to the number of ideal scores. According to Salmina et al (2017, p.43), the test item difficulty index (IK) about the form of constructed response or description is calculated using the following formula.

$$IK = \frac{\text{average score of students}}{\text{maximum score of an item}}$$

The item difficulty index was classified as: very easy, easy, moderate, difficult, or very difficult. The test instrument is said to be proportional if the higher-order mathematical thinking test instrument developed is in the category $0,30 < IK \leq 0,70$ (medium test item difficulty index).

(5) Differential power analysis

The items on the test instrument can be said to be good if the test items have a minimal discrepancy. This shows that the items have sufficient discriminatory power $0,2 \leq DB < 0,4$. Calculations to determine the number of upper and lower groups can be seen 27% of the number of students. The calculation of the differentiating power of the test items uses the following formula.

$$DB = \frac{S_A - S_B}{J_A} \quad (\text{Safitri et al, 2019})$$

Information:

S_A : total group score on an item

S_B : total score of the bottom group of an item

J_A : the number of ideal scores of an item

RESULTS AND DISCUSSION

Based on the previous description, the product developed was in the form of a test instrument to measure students' higher-order thinking skills on the material of number patterns, sequences, and series. This test instrument was developed using the ADDIE development model. This test instrument was developed with the general aim of knowing the high-order mathematical thinking skills of class VIII MTs 2 States of Medan on the subject matter of Number Patterns, Sequences, and Series and the specific purpose of this development research was to produce test instruments that were valid, reliable, level of difficulty, and good discrimination.

The validation of the test instrument was seen from the assessment of the five experts, each of whom assessed the components of the construct, language, and context of

the developed test instrument. The results of the test instrument validation are statements from experts that this test instrument is feasible to be tested with several revisions or suggestions for improvement. In addition, from the results of the test instrument validation, quantitative data was obtained which showed that the developed test instrument scored 91% with a very valid category.

After the test instrument was validated, then the test instrument was assessed through a teacher response questionnaire and a small class, each given to one teacher, and 10 students, to see the results of the responses that were used as the basis for product improvements to be developed at a later stage. From the results of the two stages, respondents were asked to respond to the available response questionnaires with the results of the responses being stated to be quantitative at 87.5% and 97.5%, both of these results indicate that the score of more than 50% of respondents gave a positive response which means the instrument test can be used without revision. After conducting the assessment stage of the test instrument by the teacher and small class trials, then the test instrument was tested on the actual class, which was 57 students.

The following is a formative evaluation that has been carried out by researchers:

1. Analysis Phase

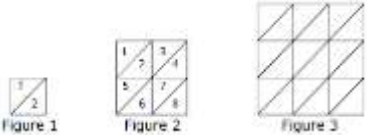
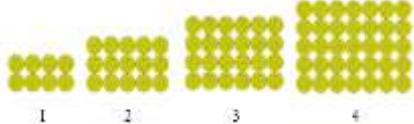
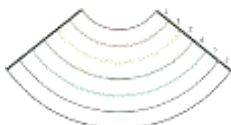
The first stage of research on the development of this test instrument is analysis, the researcher chooses a school to be used as a research location but the school is still accredited B after a discussion with the supervisor given input not to choose a school with low accreditation and it is better to choose a school that is operated/provided by the government, so the researchers chose a research school that was accredited A and operated/provided by the government, namely MTs 2 States of Medan which has become the research location.

2. Design Phase

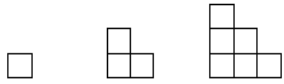


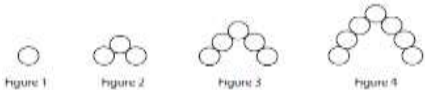
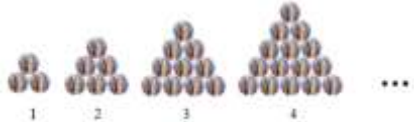
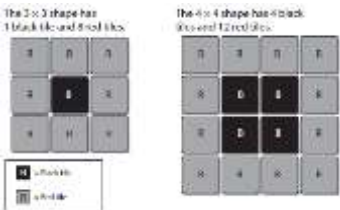

At this stage, the researcher has designed a grid of test instruments and test instruments for the subject matter of statistics where the material is even semester material, but one of the lecturers who gave suggestions for researchers and the research class teacher suggested changing the material according to the material to be studied in the odd semester, namely the material pattern of numbers, sequences, and series. So the researchers designed a grid of test instruments, test instruments, and new alternative solutions, namely on the subject matter of number patterns, sequences, and series.

The author in designing the test instrument developed by adopting the TIMSS, PISA, OSN, and HOTS question banks and then modifying them. The following is the design carried out by researchers to develop a test instrument presented in Table 1.

Table 1. Design of Each Item of Test Instruments

No	Origin	Modification	Modification Results																	
1A	<p>Source: HOTS Question Bank</p> <p>The production of batik in a batik village increases every month. The production of batik in the first month was 300 pieces and in the third month as many as 1,200 pieces. The results of the batik production for one semester are</p>	<p>Modified to ask for the blunt height of 12 seats. The reason for the modification is so that students can more easily imagine the object in question.</p>	<p>Mr. Yoyok stacked chairs of the same size that were 100 cm each. The stack height of 4 chairs is 118 cm. What is the height of the stack of 12 chairs?</p>																	
2A	<p>Source: TIMSS 1999</p> <p>The three figures below are divided into small congruent triangles</p>  <p>The sequence of figures is extended to the 7th figure. How many small triangles would be needed for Figure 7?</p>	<p>The original problem requires students to find the number of triangles from a flat square shape, modified to find the number of balls that are collected to form a rectangle.</p>	<p>Take a look at the following volleyball sets.</p>  <p>The number of volleyballs on the 30th ball is...</p>																	
3A	<p>Source: TIMSS 2003</p> <p>Betty, Frank and Darlene have just moved to Zedland. They each need to get phone service. They received the following information from the telephone company about the two different phone plans it offers. They must pay a set fee each month and there are different rates for each minute they talk. These rates depend on the time of the day or night they use the phone, and on which payment plan they choose. Both plans include time for which phone calls are free. Details of the two plans are shown in the table below.</p> <table border="1" data-bbox="183 1702 654 1948"> <thead> <tr> <th rowspan="2">Plan</th> <th rowspan="2">Monthly Fee</th> <th colspan="2">Rate per minute</th> <th rowspan="2">Free minutes per month</th> </tr> <tr> <th>Day (8 am - 6 pm)</th> <th>Night (6 pm - 8 am)</th> </tr> </thead> <tbody> <tr> <td>Plan A</td> <td>20 zeds</td> <td>3 zeds</td> <td>1 zed</td> <td>180</td> </tr> <tr> <td>Plan B</td> <td>15 zeds</td> <td>2 zeds</td> <td>2 zeds</td> <td>120</td> </tr> </tbody> </table>	Plan	Monthly Fee	Rate per minute		Free minutes per month	Day (8 am - 6 pm)	Night (6 pm - 8 am)	Plan A	20 zeds	3 zeds	1 zed	180	Plan B	15 zeds	2 zeds	2 zeds	120	<p>The problem of origin requires students to compare the cheapest prices for telephone costs, modified to design equations to find out the cheapest ticket prices. The reason for the modification, besides being required to compare, at the evaluation level must meet the ability to analyze and design. The design required in the matter of modification is to design an equation of the existing problem.</p>	<p>OSIM, a madrasah, holds an art performance for charity which is open to the general public. The committee chose a venue in the form of a theater where the audience seating was in the form of a circular sector consisting of six rows. If the first row has 25 seats, the second row has 35 seats, the third row has 50 seats, the fourth row has 70 seats, and so on. The number of spectator seats in each row forms a certain row pattern.</p>  <p>If the ticket price for the first row is the most expensive ticket and the difference in ticket prices between two adjacent rows is IDR 10,000, assuming all audience seats are fully occupied. What is the cheapest ticket price so that the committee can get an income of Rp. 22,500,000?</p>
Plan	Monthly Fee			Rate per minute			Free minutes per month													
		Day (8 am - 6 pm)	Night (6 pm - 8 am)																	
Plan A	20 zeds	3 zeds	1 zed	180																
Plan B	15 zeds	2 zeds	2 zeds	120																
4A	<p>Source: TIMSS 2003</p>	<p>The original question asked for the 10th pattern,</p>	<p>Each of the following triangles is made up of 3 sticks. Considering the</p>																	

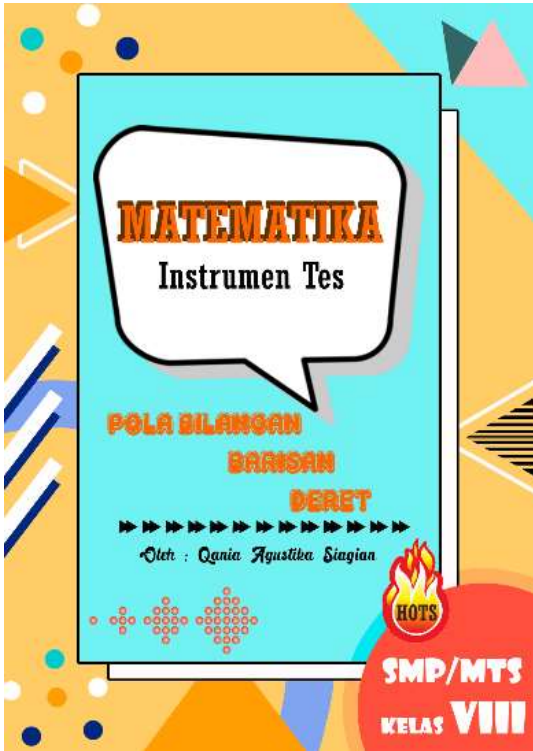

	<p>Matchsticks are arranged as shown in the figures.</p> <p>If the pattern is continued, how many matchsticks would be used to make Figure 100?</p>	<p>modified to 100th and modified the arrangement of the sticks.</p>	<p>following pattern, how many sticks are in the 100th pattern?</p>
5A	<p>Source: OSN SMP If value $100B = 100^2 + 99^2 - 98^2 - 97^2 + \dots + 4^2 + 3^2 - 2^2 - 1^2$, then the value of B is ...</p>	<p>The problem of the origin of the numbers calculated from 1 to 100, modified to be from 1 to 2011.</p>	<p>$1^2 - 2^2 + 3^2 - 4^2 + 5^2 - \dots - 2010^2 + 2011^2$ The sum of the following numbers is...</p>
1B	<p>Source: HOTS Question Bank A private employee receives a salary in the first year of Rp. 3,000,000. Every year the employee's salary increases. If the amount of money received by the employee for 10 years is RP.52,500,000, then the increase in the employee's salary each month is</p>	<p>The question of asking for a month's salary increase was modified to ask when the 40% of the bonus received was from the first-year bonus. The reason for the modification is because if you only ask for a month's salary increase, it does not require the ability to analyze.</p>	<p>Mr. Tarno is an employee of PT. Jaya Abadi who received the first annual bonus of Rp. 3,000,000. Every year the bonus received will increase by Rp. 200,000. In which year did Mr. Tarno's bonus increase by 40% from the first year's bonus?</p>
2B	<p>Sumber: TIMSS 1999 If the pattern on the grid below was continued, what letter would identify the orientation of the tile in the cell labeled X?</p>	<p>The original problem predicts the pattern that is at x, modified to predict the next pattern from the previous pattern. The reason for the modification is that it requires the ability to make hypotheses in 2 parts, the first part is the big pictures and the second part is the small picture.</p>	<p>Take a look at the following pattern!</p> <p>How is the image model in the 6th pattern?</p>
3B	<p>Source: HOTS Question Bank A private employee receives a salary in the first year of Rp. 3,000,000. Every year the employee's salary increases. If the amount of money received by the employee for 10 years is RP.52,500,000, then the increase in the employee's salary each month is</p>	<p>The original question asked for a monthly salary increase, modified to compose, and correct the hypothesis being asked. The reason for the modification is to meet the demands of ability at the level of evaluation.</p>	<p>Mr. Hadid is an employee at an insurance company who is 40 years old. This year, he became a manager and he got a salary increase of Rp. 750.000,- annually so that this year he got a salary of Rp. 15.75 million, - per month. Next year his salary will increase again to Rp. 16,500,000 per month. If the retirement limit at the insurance company is 60 years and it is assumed that Mr. Hadid will serve as manager until he retires. Has Mr. Hadid ever received a minimum salary of IDR 32,000,000 per month? If so at what age did he get it?</p>
4B	<p>Source: HOTS Question Bank At the beginning of every month, the father divides the amount of money among his 5 children. The money to be distributed consists of two</p>	<p>Regarding the origin of asking the money that the father distributed to his son, the question was modified to how much</p>	<p>Father divided the allowance as much as Rp. 300,000 to their 4 children. If the difference in money received by 2 children who are close in age is Rp. 15,000. If the youngest received the</p>



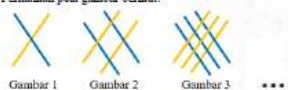




	<p>thousand bills. The first child gets 48 pieces and the second child gets half of the first child, the third child gets half of the second child, and so on. The amount of money that father distributed was....</p>	<p>money one of the children received. The reason for the modification in order to meet the demands of ability at the level of evaluating is to make hypotheses and test them.</p>	<p>least, how much money did the third child receive?</p>
<p>5B</p>	<p>Source: PISA 2014 Rohman made a staircase pattern using several squares. The following are some of the steps he goes through.</p>  <p>As shown in the figure, he used one square in Stage 1, three squares in Stage 2, and six squares for Stage 3. How many squares did he use for stage four?</p>	<p>The original problem shows a pattern by arranging a square into a ladder, modified into a circle bounded by a red line.</p>	<p>By paying attention to the circles bordered by red lines.</p>  <p>How many circles are in the 100th pattern?</p>
<p>1C</p>	<p>Source: HOTS Question Bank The production of batik in a batik village increases every month. The production of batik in the first month was 300 pieces and in the third month as many as 1,200 pieces. The results of the batik production for one semester are</p>	<p>The question of origin asking for production results for one semester, was modified to ask for a lot of production in the 8th month.</p>	<p>A factory in the first month produces 6000 units of goods and the total production for one semester is 42,000 units, the number of production in the eighth month is</p>
<p>2C</p>	<p>Source: TIMSS 2007 Posters for two sports clubs that rent bikes are shown below.</p> 	<p>The problem of origin is the cheapest cost between 2 bicycle rental shops, modified to the largest salary obtained from the 2 choices.</p>	<p>Mr. Wahyu is contracted to work for a company for 7 days. Before starting work, he was asked to choose between options (1) being given a salary of Rp. 75,000,- per day for a week, or option (2) being given a salary of Rp. 10,000, - on the first day and doubled every day for a week. Which is the best choice for Mr. Wahyu to choose?</p>
<p>3C</p>	<p>Source: TIMSS 2011</p>  <p>If there were a Figure 5, how many circles would it have?</p>	<p>Modify the arrangement of marbles and modify the question on the number of terms.</p>	<p>Consider the following set of marbles.</p>  <p>How many marbles are there in pattern 50?</p>
<p>4C</p>	<p>Source: TIMSS 2011</p>  <p>The table below shows the number</p>	<p>Modify the number of black tiles and red tiles in each square.</p>	<p>Look at the following picture.</p>  <p>How many red candies are in the tenth picture?</p>

	of tiles for the first three shapes Pat made. Pat continued making shapes using this pattern. Complete the table for the 6 x 6 and 7 x 7 shapes.		
5C	Source: HOTS Question Bank A cake is in the shape of a circle with a diameter of 25 cm. The cake will be cut into 10 square pieces. The central angles of the 10 pieces of the cake form an arithmetic sequence. If the size of the corner of the smallest piece of cake is $\frac{1}{6}$ of the size of the center of the largest piece of cake, the area of the largest piece of cake is	The original question asked for the area of the largest piece of cake, modified to ask for the highest number of shellfish in one cup. The reason for the modification to meet the demands of ability at the level of creation is to consider the incoming information.	One hundred marbles will be placed in 10 different cups. Each cup contains a different number of marbles. For the number of marbles in each cup to form an arithmetic series, what is the maximum number of marbles that can be placed in one cup?

To make this test instrument more attractive to present, the researcher designed the cover of the test instrument, instructions for use, and the contents of the test instrument. The results of the test instrument design are presented in Table 2.:

Table 2. Initial Design of Test Instruments

Cover	Question Work Instructions
	

Contents of Test Instruments	
<h3 style="color: #00796b;">Question A</h3> <p>SOAL A</p> <ol style="list-style-type: none"> Pak Yoyok memupuk kursi berukuran sama yang tingginya masing-masing 100 cm. Tinggi tumpukan 4 kursi adalah 118 cm. Berapakah tinggi tumpukan 13 kursi? Tiga buah bilangan mengikuti barisan aritmatika berjumlah 12 dan merupakan suku-suku deret aritmatika. Jika bilangan ketiga ditambah 2, maka akan membentuk barisan geometri. Berapakah hasil kali ketiga bilangan tersebut? OSDM suatu madrasah mengadakan pentas seni untuk amal yang terbuka untuk masyarakat umum. Panitia memilih tempat berupa gedung pertunjukan yang tempat duduk penontonnya berbentuk sektor lingkaran terdiri dari enam baris. Jika pada baris pertama terdapat 25 kursi, baris kedua 35 kursi, baris ketiga, 50 kursi, baris keempat 70 kursi, dan seterusnya. Banyaknya kursi penonton pada masing-masing baris membentuk pola barisan terapan.  <p>Apabila harga tiket baris pertama adalah tiket yang paling mahal dan selisih harga tiket antara dua baris yang berdekatan adalah Rp10.000,- dengan asumsi seluruh kursi penonton terisi penuh. Berapakah harga tiket yang paling murah agar panitia memperoleh pemasukan sebesar Rp22.500.000,-?</p> <ol style="list-style-type: none"> Masing-masing segitiga berikut terbentuk dari 3 stik. Dengan memperhatikan pola berikut, berapakah banyaknya stik pada pola ke-100?  <ol style="list-style-type: none"> Suku keempat, ketujuh, kesepuluh, dan ke-1010 suatu barisan aritmatika berturut-turut adalah t, t^2, $t + t^2$, dan 2018. Apakah hasil dari suku ke-100 dikurangi suku ke-10 adalah 180? Buktikanlah! 	<h3 style="color: #00796b;">Question B</h3> <p>SOAL B</p> <ol style="list-style-type: none"> Pak Tarno merupakan seorang karyawan PT. Jaya Abadi yang menerima bonus tahunan pertama sebesar Rp.3.000.000. Setiap tahun bonus yang diterima akan naik Rp.200.000. Pada tahun ke berapa bonus yang diterima Pak Tarno naik sebesar 40% dari bonus tahun pertama? Perhatikan pola gambar berikut!  <p>Diketahui gambar ke-1 membentuk 4 sudut, gambar ke-2 membentuk 16 sudut, dan gambar ke-3 membentuk 36 sudut. Berapa banyak sudut pada gambar kesepuluh?</p> <ol style="list-style-type: none"> Pak Hadid adalah seorang karyawan di sebuah perusahaan asuransi yang berumur 40 tahun. Tahun ini, ia menjadi manajer dan ia mendapat kenaikan gaji Rp750.000,- setiap tahunnya sehingga dalam tahun ini ia mendapat gaji Rp15.750.000,- per bulan. Tahun depan gajinya naik lagi menjadi Rp16.500.000,- per bulan. Apabila batas pensiun di perusahaan asuransi adalah 60 tahun dan diasumsikan Pak Hadid menjabat sebagai manajer sampai ia pensiun. Apakah Pak Hadid pernah mendapat gaji minimal sebesar Rp32.000.000,- tiap bulannya? Jika iya pada usia berapa ia mendapatkannya? Ayah membagi uang jajan sebanyak Rp.300.000 kepada 4 orang anaknya. Jika selisih uang yang diterima 2 anak yang usianya berdekatan adalah Rp.15.000. Apabila si bungsu menerima yang paling sedikit, maka berapakah uang yang diterima anak ketiga? Dengan memperhatikan lingkaran-lingkaran yang dibatasi garis merah.  <p>Berapakah banyaknya lingkaran pada pola ke-100?</p>
<h3 style="color: #00796b;">Question C</h3> <p>SOAL C</p> <ol style="list-style-type: none"> Segitiga Sierpinski adalah segitiga yang berwarna coklat pada gambar di bawah ini. Segitiga itu membagi dirinya menjadi bentuk yang sama dan ukurannya yang terus berubah.  <p>Berapakah banyaknya segitiga Sierpinski pada tahap ke-10?</p> <ol style="list-style-type: none"> Pak Wahyu dikontrak untuk bekerja pada suatu perusahaan selama 7 hari. Sebelum bekerja, ia diminta memilih antara pilihan (1) diberi gaji sebesar Rp.75.000,- per hari selama seminggu, atau pilihan (2) diberikan gaji sebesar Rp.10.000,- pada hari pertama dan bertumbuh dua kali lipat tiap harinya selama seminggu. Manakah pilihan terbaik yang harus dipilih Pak Wahyu? Perhatikan kumpulan kelereng berikut.  <p>Berapakah banyaknya kelereng pada pola ke-50?</p> <ol style="list-style-type: none"> Perhatikan gambar berikut.  <p>Berapa banyak permen merah pada gambar kesepuluh?</p> <ol style="list-style-type: none"> Seratus butir kelereng akan diletakkan pada 10 cangkir berbeda. Tiap-tiap cangkir berisi sejumlah kelereng yang berbeda. Agar jumlah kelereng pada tiap-tiap cangkir membentuk deret aritmatika berapakah jumlah kelereng terbanyak yang dapat diletakkan pada salah satu cangkir? 	

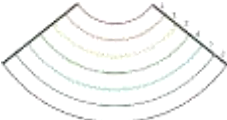

3. Development Phase




At this stage, the researchers made improvements to the quality of the test instruments through a validation process according to the directions of the validators and small class test response questionnaire analysis. From the results of the expert validation sheets, there are several questions that are recommended to be improved, such as in numbers 2A, 4A, 5A, 1B, 2B, 3B, 1C, 3C, and 4C. From the results of the student response

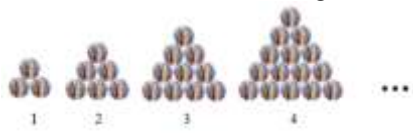
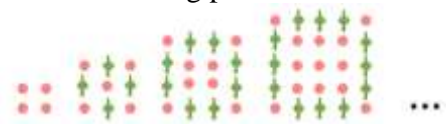
questionnaire, students were asked to color the pictures that were not colored, so the writer corrected the test by coloring the pictures that were not colored.

The results of the test instruments that have been validated by experts and the reasons for each test item can be categorized as HOTS questions which are presented in Table 3. below:

Table 3. HOTS Indicators Required in Each Test Item

No	Question	HOTS Indicator	Information
1A	Mr. Yoyok stacked chairs of the same size that were 100 cm each. The stack height of 4 chairs is 118 cm. What is the height of the stack of 12 chairs?	C4	Analyzing that the problem has a difference is always constant, which is called an arithmetic sequence. Formulate the problem that the same sized chair whose height is 100 cm is the first term.
2A	Three numbers following an arithmetic sequence add up to 12 and are terms of an arithmetic series. If the third number is added by 2, it will form a geometric sequence. What is the product of these three numbers?	C4	Correlate between arithmetic sequences and geometric sequences.
3A	OSIM, a madrasah, holds an art performance for charity which is open to the general public. The committee chose a venue in the form of a theater where the audience seating was in the form of a circular sector consisting of six rows. If the first row has 25 seats, the second row has 35 seats, the third row has 50 seats, the fourth row has 70 seats, and so on. The number of spectator seats in each row forms a certain row pattern.  If the ticket price for the first row is the most expensive ticket and the difference in ticket prices between two adjacent rows is IDR 10,000, assuming all audience seats are fully occupied. What is the cheapest ticket price so that the committee can get an income of Rp. 22,500,000?	C5	Analyze the problem that the number of seats between rows has a pattern. Designing an appropriate equation to find out the cheapest ticket prices on the condition that the first line ticket price is the most expensive and there is a difference in ticket prices between two adjacent rows.
4A	Each of the following triangles is made up of 3 sticks. Considering the following pattern, how many sticks are in the 100th pattern? 	C6	Predict/weigh/find patterns formed from pictures.
5A	The fourth, seventh, tenth, and 1010th terms of an arithmetic sequence are $t, t^2, t + t^2$, dan 2018. Is the result of the 100th term minus the 10th term 180? Prove it!	C6	Verify the information entered. Is it true that the result of the 100th term minus the 10th term is 180?

1B	<p>Mr. Tarno is an employee of PT. Jaya Abadi who received the first annual bonus of Rp. 3,000,000. Every year the bonus received will increase by Rp. 200,000. In which year did Mr. Tarno's bonus increase by 40% from the first year's bonus?</p>	C4	Find other information such as the value of U_n that can be searched from the existing information.
2B	<p>Take a look at the following pattern!</p>  <p>It is known that the 1st image forms 4 angles, the 2nd image forms 16 angles, and the 3rd image forms 36 angles. How many angles are in the tenth picture?</p>	C5	Design a pattern from the picture.
3B	<p>Mr. Hadid is an employee at an insurance company who is 40 years old. This year, he became a manager and he got a salary increase of Rp. 750.000,- annually so that this year he got a salary of Rp. 15.75 million, - per month. Next year his salary will increase again to Rp. 16,500,000 per month. If the retirement limit at the insurance company is 60 years and it is assumed that Mr. Hadid will serve as manager until he retires. Has Mr. Hadid ever received a minimum salary of IDR 32,000,000 per month? If so at what age did he get it?</p>	C5	Develop hypotheses and correct hypotheses.
4B	<p>Father divided the allowance as much as Rp. 300,000 to their 4 children. If the difference in money received by 2 children who are close in age is Rp. 15,000. If the youngest received the least, how much money did the third child receive?</p>	C5	Designing known information, such as the statement that the youngest received the least, said the youngest means that the child is the 4 th or 4 th tribe.
5B	<p>By paying attention to the circles bordered by red lines.</p>  <p>How many circles are in the 100th pattern?</p>	C6	Predict/weigh/find patterns formed from pictures.
1C	<p>The Sierpinski triangle is the brown triangle in the image below. The triangle divides itself into equal shapes and constantly changing sizes.</p>  <p>How many Sierpinski triangles are there in step 10?</p>	C4	Find the pattern in the Sierpinski triangle.
2C	<p>Mr. Wahyu is contracted to work for a company for 7 days. Before starting work, he</p>	C4	Compare the statement of choice (1) with choice (2).

	was asked to choose between options (1) being given a salary of Rp. 75,000,- per day for a week, or option (2) being given a salary of Rp. 10,000, - on the first day and doubled every day for a week. Which is the best choice for Mr. Wahyu to choose?		
3C	Consider the following set of marbles.  How many marbles are there in pattern 50?	C5	Predict the pattern of the set of marbles.
4C	Look at the following picture.  How many red candies are in the tenth picture?	C6	Predict/weigh/find patterns formed from pictures.
5C	One hundred marbles will be placed in 10 different cups. Each cup contains a different number of marbles. For the number of marbles in each cup to form an arithmetic series, what is the maximum number of marbles that can be placed in one cup?	C6	Weighing the incoming information.

4. Implementation Phase

At this stage, the researcher conducts a summative evaluation, namely the assessment of test instruments to make decisions about the sustainability of a product. Summative evaluation is carried out by knowing the validity and reliability of the test items, the level of difficulty, and the distinguishing power of the developed test instrument.

Based on the results of this study, all items A on this test instrument is said to be valid. The criteria for item numbers 1, 2, 3, and 5 have "High" validity, while item 4 has "Very High" validity. For all items B on this test instrument is said to be valid. For the criteria items, numbers 2 and 5 have "Enough" validity, 1 and 4 have "High" validity, while item number 3 has "Very High" validity. For all items C on this test instrument is said to be valid. The criteria for items number 1, 2, 4, and 5 have "High" validity, while item number 3 has "Enough" validity.

Judging from the reliability of the test items, item A on this test instrument has "High" reliability of 0.76. For item B on this test instrument has "High" reliability of 0.67. Meanwhile, item C on this test instrument has "High" reliability of 0.69. Judging from the analysis of the level of difficulty of the items, for all questions A, B, and C, all three have a "Medium" difficulty index.

Judging from the analysis of discriminating power, for question A in questions number 1, 2, 3, 4, and 5 have a different power of "Good". For question B on questions number 2, 4, and 5 it has a "good enough" difference power and for questions number 1 and 3 it has a "good" distinguishing power. For question C on questions number 1, 2, 3, 4, and 5 it has a "Good" distinction.

Table 4. Students' Mathematical High Order Thinking Skill (HOTS) Test Results

Score	Frequency	Percentage (%)	Category
$90 \leq \text{Score} \leq 100$	8	14,04	Very high
$75 \leq \text{Score} < 90$	19	33,33	High
$55 \leq \text{Score} < 75$	12	21,05	Enough
$40 \leq \text{Score} < 55$	11	19,3	Less
Score < 40	7	12,28	Very less
Number of Subjects	57	100	
Average score	65,09		Enough

Based on the table above, it is known that the students' higher order mathematical thinking skills were found that from 57 subjects there were 8 students (14.04%) included in the category of having higher-order mathematical thinking skills in the very high category, 19 students (33.33%) having higher order thinking skills. high mathematical ability in the high category, 12 students (21.05%) have high order mathematical thinking skills in the sufficient category, 11 students (19.3%) have high order mathematical thinking skills in the poor category, and 7 students (12.28%) have high order mathematical thinking skills in the very poor category. So that the average high order mathematical thinking ability of students is 65.09% with sufficient category. The students' higher order thinking ability is still in the sufficient category because the question is a non-routine question when learning in class with the teacher so far. This was evidenced by the results of one student response questionnaire which said that this test instrument was better for exam questions, this was because they were not familiar with answering HOTS questions.

In line with the research by Budiman and Jailani (2014) in Aisah & Pahlevi (2020, p.155), one of the causes of below-average student abilities is that students lack practice in working on contextual questions, which require reasoning, argument, and creativity skills. If teachers are able to encourage students to think at a higher level then students will have higher-order thinking skills that will be able to compete in the TIMSS and PISA assessments.

The responses of the 57 students to the given test instrument showed a positive response. The average shows 94.74%, which means that more than 50% of students give a positive response, so it can be said that it can be used without revision.

This research is also relevant to the research conducted by Martina (2017) based on the results of the research that the test instrument developed can measure students' HOTS by meeting the quality criteria of the test instrument. These criteria include validity, reliability, level of difficulty, and distinguishing power. Also in line with Nurmadinah's opinion (2017) based on the results of her research that the test instrument developed if it is declared valid, has good reliability, a decent level of difficulty, and not bad discriminating power, the test instrument can measure students' HOTS.

This proves that the test instrument developed is truly appropriate and can be used by a wider range of schools. By fulfilling the stages of testing the validity, reliability, level of difficulty, and discriminating power. So the stage of developing this test instrument has ended.

CONCLUSION

Based on the results of the previous research and discussion, it can be concluded that 15 questions have been successfully developed which are declared valid, this is the average score of the test instrument assessment by experts of 91% and the validity of each test item is declared "valid". The reliability of the test instrument for each package of questions was assessed as reliable with a score of 0.76, 0.67, and 0.69 in the high category. In the analysis of the level of difficulty, it appears that the results are in the easy and medium categories. In the analysis of discriminatory power, it appears that the results are quite good and good. The average result of students' higher-order mathematical thinking skills is 65.09% with sufficient category.

The researcher gives the advice to further researchers, this question can be used as material for consideration to examine more deeply problems of higher-order mathematical thinking skills for other materials so that students are familiar with questions that meet the criteria for higher-order mathematical thinking skills. And the teacher should have provided training in the form of HOTS questions continuously so that students are trained and accustomed to working on HOTS questions so that students can support their ability to think at higher levels.

THANK-YOU NOTE

The authors would like to thank Dr. Izwita Dewi, M.Pd., Mr. Mangaratua M. Simanjorang, M.Pd., Ph.D., and Mrs. Glory Indira D. Purba, S.Si., M.Pd., as lecturers in the Mathematics Education Study Program at the University Medan State who has provided advice and direction during the research process so that this research can run smoothly.

REFERENCES

- Aisah, S., & Pahlevi, T. (2020). Pengembangan Instrumen Penilaian Higher Order Thinking Skills (HOTS) Pada Mata Pelajaran Korespondensi Kelas X OTP di SMK Negeri 1 Jombang. *Jurnal Pendidikan Administrasi Perkantoran (JPAP)*, Vol. 8 (1), 146-156.
- Ardhana, T. (2017). *Keterampilan Berpikir Siswa Dalam Menyelesaikan Soal Garis Dan Sudut Berdasarkan Taksonomi Bloom Revisi*. Skripsi. Universitas Muhammadiyah Surakarta: Tidak diterbitkan.
- Arifin, Z. (2017). *Evaluasi Pembelajaran*. Bandung: PT Pemuda Rosdakarya.
- Arlianty, W. N., Febriana, B. W., Diniaty, A., & Fauzi'ah, L. (2018). Student Profile in Completing Questions Based on Cognitive Level of Bloom's Taxonomy By Anderson and Krathwohl. *AIP Conference Proceedings*, Vol. 2026 (020063), 1-6.
- Dewi, R. (2017). *Pengembangan Instrumen Tes Untuk Mengukur Kemampuan Komunikasi Matematis Siswa SMP Negeri 17 Makassar*. Skripsi. UIN Alauddin Makassar: Tidak diterbitkan.
- Fadlila, Nindy. (2019). *Pengembangan Instrumen Tes Berbasis Pendidikan Realistik Untuk Meningkatkan High Order Thinking Skill (HOTS) Siswa SMP Muhammadiyah 47 Sunggal T.A. 2018/2019*. Skripsi. Universitas Negeri Medan: Tidak diterbitkan.
- Febiana, M. V. (2019). *Pengembangan Soal HOTS Materi Luas Bangun Datar Dan Volume Bangun Ruang Untuk Siswa Kelas V SD*. Skripsi. Santa Dharma University Yogyakarta: Tidak diterbitkan.
- Febri E. (2017). *Pedoman Metodologi Penelitian*. Sidoarjo: Zifatama Jawa.
- Hadi, S. (2018). *Pendidikan Matematika Realistik dan Implementasinya*. Banjarmasin: Tulip.
- Hajar, I., Lufri, & Fauzan A. (2021). Validity Level Analysis Model Pjbl For Biology Learning. *Journal of Physics: Conference Series*, Vol. 1940 (012125), 1-7.
- Jannah, M., Abdi, A. W., & Harun, M. Y. (2017). Hubungan Pemanfaatan Internet Sebagai Sumber Belajar Dengan Hasil Belajar Geografi Siswa Kelas X Lab School Banda Aceh. *Jurnal Ilmiah Mahasiswa Pendidikan Geografi FKIP Unsyiah*, Vol. 2 (2), 78-91.
- Khodaria, S., Maharani, A., & Sulaiman H. (2019). The Analysis Of Item Problems In High School Mathematics Textbook In Indonesia (2016 Revision Edition) Reviewed From The Cognitive Aspect Of TIMSS. *Indonesian Journal of Learning and Instruction*, Vol. 2 (1), 65-70.
- Kusuma, M. D., Rosidin, U., Abdurrahman, & Suyatna, A. (2017). The Development of Higher Order Thinking Skill (HOTS) Instrument Assessment In Physics Study. *IOSR Journal of Research & Method in Education*, Vol. 7 (1), 26-32.
-

- Malik, R. S. (2018). Educational Challenges In 21st Century And Sustainable Developmen. *Journal of Sustainable Development Education and Research*, Vol 2 (1), 9-20.
- Martina. (2017). *Pengembangan Instrumen Tes Higher Order Thinking Skill (HOTS) Pokok Bahasan Sistem Persamaan Linear Dua Variabel Dan Teorema Pythagoras Kelas VIII SMP Citra Samata Kab. Gowa*. Skripsi. Alauddin State Islamic University Makassar: *Tidak diterbitkan*.
- Motlhabane, Abraham. (2017). Unpacking The South African Physicsexamination Questions According To Blooms' Revised Taxonomy. *Journal of Baltic Science Education*. Vol, 16 (6), 919-931.
- Nurmadinah. (2017). *Pengembangan Instrumen Tes Higher Order Thinking Skill (HOTS) Pokok Bahasan Operasi Hitung Bentuk Aljabar Serta Persamaan Dan Pertidaksamaan Linear Satu Variabel Kelas VII MTs Guppi Samata*. Skripsi. Alauddin State Islamic University Makassar: *Tidak diterbitkan*.
- OECD. (2019). *PISA 2018 Result In Focus*. Paris: OECD.
- Rapih, S., & Sutaryadi. (2018). Perspektif Guru Sekolah Dasar Terhadap High Order Thinking Skills (HOTS): Pemahaman, Penerapan, dan Hambatan. *Jurnal Pendidikan Dasar dan Pembelajaran*, Vol. 8 (1), 76-87.
- Safitri, E., Sofiyani, Ramadhani D., & Sukirno. (2019). Analisis Soal Ujian Matematika Semester Ganjil Pada SD Negeri 10 Langsa. *Journal of Basic Education Studies*, Vol 2 (1), 89-99.
- Salmina, M., & Adyansyah F. (2017). Analisis Kualitas Soal Ujian Matematika Semester Genap Kelas Xi Sma Inshafuddin Kota Banda Aceh. *Numeracy Journal*, Vol. 4 (1), 37-47.
- Sugiyono. (2017). *Metode Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Wulan, S., Jufrida, & Pathoni, H. (2017). Pengembangan Modul Elektronik Berbasis 3D Pageflip Professional pada Materi Konsep Dasar Fisika Inti dan Struktur Inti Mata Kuliah Fisika Atom dan Inti. *Jurnal EduFisika*, Vol. 2 (1), 38-50.
-