Journal for Lesson and Learning Studies
Volume 4, Number 2, 2021 pp. 271-280
P-ISSN: 2615-6148, E-ISSN : 2615-7330
Open Access: https://ejournal.undiksha.ac.id/index.php/JLLS

Hot Based Assessment Instrument to Assess Math Learning Achievement

Ni Nyoman Tri Rapianingsih[1]*, I Made Suarjane[2], I Ketut Dibia[3]
[1,2,3] PGSD, Universitas Pendidikan Ganesha, Singaraja,Indonesia

ARTICLE INFO

Article history:
Received March 07, 2021
Revised March 10, 2021
Accepted June 09, 2021
Available online July 25, 2021

Kata Kunci:
HOTS, Instrumen Asesmen Matematika, Model 4D

Keywords:
HOTS, Mathematical Assessment Instrument, 4D Model

ABSTRAK

Minimnya ketersediaan instrumen penilaian untuk mengukur kemampuan kognitif tingkat kemampuan C4-C6, kurangnya pemberian soal latihan tingkat C4-C6 dalam pembelajaran Matematika mengakibatkan penilaian yang dilakukan kurang optimal. Tujuan penelitian adalah untuk menguji kualitas instrumen penilaian hasil belajar matematika berorientasi HOTS. Penelitian ini merupakan penelitian pengembangan model 4D oleh Thiagarajan. Tahapan tersebut adalah: define, design, develop, dan disebarluaskan. Penelitian yang dilakukan baru sampai pada tahap pengembangan karena keterbatasan waktu, biaya, dan tenaga. Subjek penelitian adalah kisi-kisi dan tes hasil belajar pilihan ganda sebanyak 20 soal. Metode pengumpulan data adalah metode tes dan non tes. Instrumen pengumpulan data berupa lembar angket juri dan tes hasil belajar. Metode analisis data validitas isi menggunakan rumus Gregory, dan validitas item menggunakan rumus Product moment, daya pembeda, tingkat kesukaran, kualitas distraktor, dan reliabilitas menggunakan rumus KR-20. Hasil analisis data diperoleh uji validitas isi 1,00 dengan kriteria sangat tinggi. Memiliki daya pembeda 16 item dengan kriteria "cukup baik", 2 dengan kriteria "baik". Memiliki indeks kesukaran 1 butir soal dengan kriteria "sulit", 14 butir soal dengan kriteria "sedang", dan tiga butir soal dengan kriteria "mudah". Memiliki kualitas mengganggu P > 5% sebanyak 53, P < 5% sebanyak 7. Dengan uji reliabilitas 0,93 kategori sangat tinggi. Berdasarkan hasil penelitian ini, pengembangan instrumen penilaian hasil belajar matematika berorientasi HOTS adalah valid, reliabel, dan layak digunakan.

ABSTRACT

The lack of availability of assessment instruments to measure cognitive ability at the C4-C6 ability level, the lack of giving practice questions at the C4-C6 level in Mathematics learning resulted in the assessment being carried out less than optimal. The research objective is to test the quality of the HOTS-oriented mathematics learning outcomes assessment instrument. This research is research on the development of 4D models by Thiagarajan. The stages are: define, design, develop, and disseminate. The research carried out only reached the development stage due to limited time, cost, and manpower. The research subjects were lattices and learning outcomes tests in multiple-choice as many as 20 questions. Data collection methods are test and non-test methods. The data collection instruments were judges' questionnaire sheets and learning outcomes tests. Content validity data analysis method used the Gregory formula, and item validity used the Product moment formula, discriminatory power, difficulty level, distractor quality, and reliability uses the KR-20 formula. The results of data analysis obtained a content validity test of 1.00, very high criteria. It has a distinguishing power of 16 items with the criteria of "good enough," 2 with the criteria of "good." It has a difficulty index of 1 item with "difficult" criteria, 14 items with "medium" criteria, and three items with "easy" criteria. Has a distracting quality of P > 5% as many as 53, P < 5% as many as 7. With a reliability test of 0.93, the category is very high. Based on the results of this study, the development of the HOTS-oriented mathematics learning outcome assessment instrument is valid, reliable, and feasible to use.

1. INTRODUCTION

The assessment instrument given to students is still at the C1-C3 level, impacting student learning outcomes who have not been able to solve more complex problems. Even though the Industrial Revolution 4.0 has spread to every aspect of life, one of which is education. To face the 4.0 industrial

Corresponding author
*E-mail addresses: trirapianingsih1998@gmail.com (Ni Nyoman Tri Rapianingsih)

revolution, education is directed to develop 21st-century skills (Ginting & Haryati, 2012). 21st-century skills, commonly called 4Cs, consist of critical thinking, collaboration, creativity, and communication (Ahmad et al., 2018; Antara et al., 2020; Trisnawati & Sari, 2018). Education in Indonesia is currently making various efforts to improve one of them by implementing the 2013 curriculum. The 2013 curriculum demands higher-order thinking skills since the elementary school level. Higher-order thinking (HOTS) is part of the revised Blom Taxonomy in the form of operational verbs consisting of analyze (C4), evaluate (C5), and create (C6), which can be used in learning activities (Mislikhah, 2020; Rahayu et al., 2020). HOTS are thinking skills possessed by students more broadly who can apply concepts, analyze an idea, and even create an idea (Ahmad et al., 2018). Higher-order thinking skills are critical to be developed in students. Higher-order thinking skills that students must possess include critical thinking in solving problems, creativity, innovation, communication, collaboration, self-confidence, ability to argue (reasoning), and the ability to make decisions (Antara et al., 2020; Febrina et al., n.d.; Purnama & Nurdianingsih, 2019). HOTS needs to be applied to the learning process in the classroom, and the learning carried out provides space for students to work on HOTS-type questions. This is, of course, the teacher must master the material, learning strategies, and the ability to prepare HOTS questions so that students can develop higher-order thinking skills, one of which is in learning mathematics (Hanifah, 2019; Kristanto, 2020; Tyas et al., n.d.). HOTS learning is one of the priority subjects in mathematics at school. Through the application of HOTS learning, students can reason, problem-solve, mathematical connections, mathematical literacy, and mathematical representation so that the objectives of learning mathematics can be achieved optimally (Antara et al., 2020; Kaur Swaran Singh et al., 2019; Ricardo, 2017).

The low level of students' high-order thinking skills is caused by the lack of teacher ability in developing the HOTS instrument. In schools, they tend to test aspects of memory and do not practice higher-order thinking questions (Budiman & Jailani, 2014; Ikhsan & Handayani, 2016). In addition, it is still emphasized in learning mathematics that memorizing formulas cause students' higher-order thinking skills to not develop properly (H. Handayani, 2015; Singsh et al., 2019). Teachers give lectures and practice questions without understanding the concept of the material in-depth, which results in students being less trained to develop their reasoning abilities in solving problems and applying concepts that have been learned in real life so that students' thinking abilities do not develop optimally (Budi, 2014; Sukla & Dungsungneon, 2016). The old teacher-centered learning process is still heavily implemented by the teacher, the constructivist view that involves more students is still rarely carried out, besides that the teacher still lacks innovation in inviting students to practice answering questions (Kholis, 2017; Muslich, 2011; Wijayanti, 2014). Problems were also found based on the results of interviews that had been conducted with homeroom teachers for grade IV SD in Cluster VII, Buleleng District. There were still many problems in the preparation of assessment instruments, namely as follows: 1) teachers had difficulties in making assessment instruments, especially in mathematics lessons at levels C4-C6, 2) the lack of understanding of technology owned by the teacher has an impact on the teacher's lack of creativity in developing the questions that will be given to students, 3) students do not understand the questions given by the teacher, 4) the provision of HOTS type questions exercises that are still not given to students, and 5) the questions given to students are still at levels C1, C2, and C3. If this continues, the abilities possessed by students cannot develop towards higher-order thinking skills. Because the learning activities that occur do not train students to develop students' higher-order thinking skills so that students' high-level cognitive abilities are feeble due to learning activities carried out only to encourage students to think at a lower level (Nasrum, 2020; Putra, 2017).

Based on these problems, students' higher-order thinking skills in participating in the learning process cannot be separated from how a teacher conveys material through approaches, models, and methods in the learning process. To find out the success of a learning test, it can be done by providing a learning evaluation in the form of an assessment instrument in practice questions. Therefore, the teacher plays a role in optimizing the HOTS assessment, both in daily tests, end-of-semester assessments, and school exams, to train and determine the categories of higher-order thinking skills that students have (Kusainun, 2020; Widana, 2017). Assessment instruments need to be developed to measure student learning outcomes that include learning outcomes in the cognitive, affective, and psychomotor domains (Wirayasa et al., 2020; Yuliandini et al., 2019). In general, there are two ways to collect student data after participating in the learning process, namely the test and non-test techniques according to the aspect you want to assess. The test technique is used to measure students who have participated in the learning process in terms of knowledge and skills.In contrast, non-test techniques are used to collect information/data of students who have participated in the learning process related to aspects of personality, attitudes, characteristics, and skills of students (Desilva et al., 2020; Lestari & Subiyanto, Hadisaputro Murbangun, 2015; Uno & Koni, 2012). Assessment needs to be carried out to measure the achievement of student competencies on an ongoing basis. This is the basis in the learning process to monitor progress and improve student learning outcomes (Suci, 2020). So that assessment activities are essential to be given in the learning process because they can provide constructive feedback for teachers and students, motivate students to excel, and influence learning behavior by the assessment carried out by the teacher (Dharmawati, 2016).

For students to have high-order thinking skills, there needs to be habituation through working on High Order Thinking Skill (HOTS) questions in the mathematics learning process (Astiwi et al., 2020; Nurfillaili et al., 2016; Rahayu et al., 2020). HOTS is complex thinking processability that includes parsing material, criticizing, and creating solutions to a problem faced by students (Budiarta et al., 2018; Koyan, 2012). The characteristics of HOTS questions are (1). questions that are made can measure students' high-level abilities, (2) questions that are contextual based and interesting to be worked on by students, and (3) are carried out in stages and raise new problems to be solved by students (Farhan, 2018; Hardiani, 2017). Higher-order thinking skills are measured from the answers to questions that have been made based on the revised Bloom's Taxonomy, from higher-order thinking skills like problem-solving needed in the learning process because learning oriented to higher-order thinking cannot be separated from creativity skills in solving problems (Hapsari & Kusnindyah, 2020). To develop students' higher-order thinking skills, it is necessary to develop instruments at the cognitive level C4-C6. Giving HOTS questions routinely and gradually will change the students' thinking (Fatimah & Alfath, 2019).

Research that is relevant to development research is research conducted by Mailani (2019) which shows the E-Quiz-based assessment instrument developed is in the proper category with a validity value of 86.5% with a very good category. Then the research that is conducted by Niswara et al. (2019) shows the developed test package has a reliability of 0.803 with a high-reliability category. From the research above, there are differences, namely this research develops mathematical instruments on the flat material. Based on the problems that have been described and the study of supporting theories, it is necessary to develop a HOTS-oriented mathematics learning outcome assessment instrument. This research is important to do to improve the assessment instrument that can develop higher-order thinking skills. The difference between the HOTS-oriented Mathematics learning outcome assessment instrument and the existing instruments is the test indicator which refers to the C4-C6 cognitive level. The purpose of this study is to test the quality of the HOTS-oriented mathematics learning outcomes assessment instrument. This research implies a HOTS-oriented mathematical learning outcome assessment instrument that is feasible to use as a guide in developing HOTS-oriented assessment instruments on the flat shape material for grade IV SD. The contribution of ideas related to the development of instruments in mathematics learning is that the teacher fully supports the practice to improve the quality possessed in developing and improving assessment instruments. Other research can be used as a reference for conducting similar research, which is developing to the dissemination stage so that more information is known about the effectiveness of the assessment instruments that have been developed.

2. METHOD

The type of research design is developmental research. This study developed an instrument for assessing mathematics learning outcomes with a HOTS orientation. This research model uses a 4D model that Marzuki and Seri (2017) coined. The research steps consisted of 4 stages: the definition stage, the design stage, the development stage, and without the deployment stage due to the limitations of time, cost, and energy possessed by the researcher. The defining stage is to analyze the needs of students through observation and interview techniques. The design stage is carried out by selecting the format for the components in the assessment instrument and making the instrument design by the existing format. The development stage carried out is conducting a feasibility test by testing the validity of the product to experts in the field of mathematics lessons. The 4D model was chosen because it is based on the consideration that this model is simple. The stages are systematically based on a theoretical basis and a directed development flow. The subjects of this study are the grid and the test of learning outcomes with the number of trial samples as many as 69 students. The data collection methods used in this study are the test method and the non-test method. In this study, the data required is the results of the content validity test using the non-test method in the form of a list of relevant and irrelevant lists. While the data from the test results were obtained using the test method in the form of multiple-choice as many as 20 questions. Table 1 below. The assessment instrument developed was assessed by two lecturers as experts in the field of mathematics. The validator assesses the instrument that has been developed on every aspect of the question indicator that has been designed with relevant and irrelevant values. Relevant means that the items developed match the designed test, sentence/language, and the grid. Irrelevance means that the validator assesses that the instrument developed is not by the designed test, sentences/language with the designed grid, and needs improvement. The lattice of the assessment instrument developed into a test of learning outcomes which experts/experts will then validate as Table 2.

Table 1. Method, Instrument, Data Collection Technique

Research purposes	Method of collecting data	Data Collection Instruments	Data analysis technique
To test the quality of the HOTS-oriented Mathematics learning outcomes assessment instrument for fourth-grade elementary school	1. . Test meth 2.Non-test method	1. Expert validity sheet 2. Study result test	1. Cross tabulation 2x2 2. Product moment inferential analysis technique 3. KR-20 statistical analysis technique

Table 2. Instrument Blueprint

Basic Competence	Indicator	Knowledge Aspect	Question Number
3.9 Explain and determine the perimeter and area of squares, rectangles, and triangles	3.9.9 Analyze various shapes of squares, rectangles, and triangles.	C4	1,2
	3.9.10 Synthesize problems in calculating and determining the perimeter of a square.	C5	3,4,5
	3.9.11 Solve everyday problems related to the area of a square.	C4	6,10
	3.9.12 Solve problems in calculating and determining the perimeter of a rectangle.	C4	7,9
	3.9.13 Solve everyday problems related to the area of a rectangle	C4	8,11,15
	3.9.14 Solving problems in calculating and determining the perimeter of a triangle	C4	12,14
	3.9.15 Comparing the relationship between perimeter and area of flat shapes (triangles, rectangles)	C5	13,20
	3.9.16 Designing alternative solutions to problems related to the combined area of flat shapes	C6	16
4.10 Solve problems related to the perimeter and area of squares, rectangles, and triangles	4.9.10 Synthesize everyday problems in determining the perimeter and area of flat shapes.	C5	18,19
	4.9.11 Designing alternative solutions to open problems related to the area of flat shapes.	C6	17

The data obtained were processed using quantitative descriptive analysis. Quantitative descriptive data analysis was used to determine the quality of the developed instrument in terms of validity, reliability, and discriminating power, level of difficulty, distractor quality, and reliability. Whether or not an instrument is developed is determined by its validity and reliability. Instrument validity is sought to make precise measurements in measuring what is to be measured, while reliability is the extent to which a measurement can be trusted because of its stability (Febrianawati, 2018). To find content validity using the Gregory 2x2 formula with the following mechanism: 1). Experts/experts assess the instrument with a score of 3 or 4 categories very relevant, and a score of 1 or 2 less relevant 2). The expert/expert assessment results are cross-tabulated with a matrix (2 x 2), as shown in Table 3 below (Setemen, 2018). The next step is calculating content validity. The validity level of the results of the calculation of content validity can be seen in the category of content validity coefficients according to Table 3.

Table 3. Criteria of Content Validity

Coefficient	Validity
0,80-1,00	Very high
0,60-0,79	High
0,40-0,59	Medium
0,20-0,39	Low
0,00-0,19	Very low

The instrument that has been tested will be obtained by student data. To obtain valid data, the instrument or tool used to evaluate must be valid. The correlation technique used to test the validity of the test items of the HOTS-oriented mathematics learning outcome assessment instrument is the Moment Product correlation technique. To find out the intensity of a question in terms of difficulty, a discriminatory power is needed. The discriminatory test was given to distinguish between high-ability students and low-ability students. The calculation of the different power is divided into two groups, namely small groups and large groups, with small groups using 30% of the upper group and 30% of the lower group. While the large group used 27% of the upper group and 27% of the lower group, in this study, the calculation of discriminating power used 27% of the upper group and 27% of the lower group. To obtain good quality questions and meeting validity and reliability, there is a balance between the level of difficulty of the questions (Fatimah & Alfath, 2019; Zainal, 2018). Items can be said to be good if the tests carried out by students are not too difficult and not too easy. In other words, the degree of difficulty of the test is moderate and sufficient.

The distractor has carried out its function well if the participant feels tricked into choosing the distractor as the correct answer/answer key when it is not. The quality of distractors can be seen by looking at the distribution of students' answers. A distractor is said to function properly if it is chosen by at least 5% of the test takers. To test the reliability of the HOTS-oriented mathematics learning outcome assessment instrument, the KR-20 statistical analysis technique was used. Instruments that can be tested for reliability using KR are instruments with one correct answer key (Febrianawati, 2018). Reliability is the level of constancy or consistency of measurement, can be trusted to produce a steady score, and relatively unchanged even though tested in different situations. To test the reliability of the HOTS-oriented mathematics learning outcome assessment instrument, the KR-20 technique was used.

3. RESULT AND DISCUSSION

Results

The model is used in 4D, with the first stage being definition, the second is design, the third is development, and the fourth is deployment. Based on these stages, the results form a HOTS-oriented assessment instrument on the Build Flat material. **Define stage** done to define and define learning needs. The study began with conducting observations by conducting interviews with teachers at SD Cluster VII, Buleleng District, to collect data related to students' mathematics learning outcomes and the cognitive level of the questions given to students. The final preliminary analysis is carried out by determining the needs of students by analyzing the math syllabus for even semesters. Based on the results of interviews with fourth-grade teachers, it can be seen that during the evaluation of the learning process, the instruments given to students did not refer to questions that lead students to higher-order thinking (HOTS), especially in Mathematics. The instruments given to students are still at the LOTS level from C1-C3, which results in students' thinking skills being low and have not been able to perform analysis and solve complex problems. Therefore, innovation is needed to overcome this in the form of an instrument that can train students' higher-order thinking skills in analyzing, solving problems, especially in the material for flat-building for grade IV elementary school. The results of the analysis of the curriculum used at SD Cluster VII, Buleleng District, namely the 2013 curriculum. From the analysis of student characteristics, namely, there are still indicators that have not been listed in the questions given to students. Students do not understand the questions given by the teacher, there are no students invited to practice questions. for higher-order thinking. From the analysis of the students' initial and final analysis, it can be determined the material to be developed can encourage students' higher-order thinking skills, namely the Build Flat material. The analysis that has been passed from the student and material analysis results continues with the formulation of the indicators used in the questions, which consist of C4-C5.

The **design stage** carried out after going through the analysis phase is complete. The design begins with designing the grid. The grid is arranged based on KD, KI, and indicators. After the lattice design is completed, it is continued by compiling the initial design in draft I of the HOTS-oriented mathematics learning outcome assessment instrument for flat-building materials. At the initial product development stage of this research, the instrument for assessing mathematics learning outcomes with the HOTS orientation with the material *Bangun Datar*. The instrument used in the form of multiple-choice consisting of 20 questions. The instrument developed is by the Basic Competencies in the even semester Mathematics syllabus. The instrument developed was adapted to the revised Bloom's Taxonomy in the form of operational verbs consisting of analyze (C4), evaluate (C5), and create (C6). Through the provision of these instruments will develop higher-order thinking skills in learning Mathematics. After the instrument has been developed, the validity test is carried out by giving the material expert test/judges an assessment sheet. This is done to determine the validity of the developed instrument. The validity test is carried out by providing an assessment sheet in the form of a list to material expert one and material expert 2. The results of the relevance of the instrument obtained from expert one and expert 2 are in table 4 below.

Table 4. Content Validity

Expert 1		**Expert 2**	
Relevant	**Irrelevant**	**Relevant**	**Irrelevant**
1,2,3,4,5,6,7,8,9,10 11,12,13,14,15,16,17, 1819,20	-	1,2,3,4,5,6,7,8,9,10 11,12,13,14,15,16,17, 1819,20	-

Based on the data above, it can be seen that the 20 questions developed were stated to be relevant. Expert 1 and expert 2 stated that the developed instrument was relevant, which means that the grid and the developed learning outcomes test instrument were appropriate for the language, sentences, and material. The assessment obtained was then analyzed for content validity using the Gregory formula. The calculation of the coefficient of content validity of the developed instrument product is 1.00 seen from the content validity category, which is in the very good category. This shows that based on the assessment carried out by expert judges, the instrument developed has been based on the theory of instrument preparation and has a match between indicators and questions, grammar, and constructs. The next stage is **testing**. The trials carried out in class V are 69 students; trials are carried out to determine whether the items developed have met good quality questions and are suitable for use. The calculation results after the trial were conducted by looking for the results of the validity of the test items as many as 20 multiple-choice questions. The data from the item validity test results using the moment product correlation formula assisted by the Microsoft Office Excel 2010 application. It shows that 18 items are declared valid from the 20 items tested. 2 items were declared invalid and unfit to be used as an assessment instrument. 18 Items that are declared valid indication that the instrument developed has accuracy and accuracy in carrying out its measuring function to measure students' higher-order thinking skills. In contrast, the 2 test items were declared invalid because the two test items did not have the accuracy and accuracy in carrying out their measuring function to measure students' higher-order thinking skills.

The next step is to analyze the differential power. The results of the analysis of the items' differentiating power obtained the results that 16 items were on the "Good enough" distinction criteria and two items were on the "Good" criteria. This shows that the developed instrument has a fairly good ability in distinguishing students who have mastered the material from students who have not mastered the material of flat shapes. Two items are in a good category, which means that it is a good question to distinguish the abilities of smart students and students who are less intelligent. In comparison, the 14 items are in the good enough category to distinguish the abilities of smart students and students who are less intelligent. The next stage is to analyze the level of test difficulty. The results of the analysis of the level of difficulty of the items showed that 1 item was in the "difficult" category of difficulty, 14 items were in the "medium" difficulty category, and three items were in the "easy" category. This can be interpreted that the instrument developed has a quality level of difficulty that is neither difficult nor easy. One item in the difficult category means that the question is difficult to work on for students who have the ability of smart students. Fourteen items are in the medium category, meaning that fast learners and slow learners can answer the questions. Three questions are in the easy category, which means students who have low abilities can answer correctly.

The next stage is to analyze the quality of distractors. The item analysis results show that the distractors are said to be good if at least 5% of the total number of students are selected. In the calculation results above, the distractors who have P > 5% are 53 and P 5% are 7. 53 distractors have functioned well, meaning that the distractors can confuse students in determining the correct answer. Meanwhile, the seven distractors did not carry out their function as distractors properly, which means that the distractors could not make students mislead between the answer key and the non-answer key. This can be interpreted that the quality of the detractor developed can function well. The next stage is reliability analysis. The reliability coefficient obtained from the item analysis results is 0.93, with a very high category. This shows that the instrument product developed has high test criteria (reliable) to be tested at any time with fixed or relatively fixed results on equal respondents. Based on the results of expert validation, testing, data analysis, and improvements made, it can be seen that the developed instrument has met the valid and reliable criteria, and the quality of the items is good so that the developed instrument can be used as an evaluation tool or an accurate learning assessment on the cognitive aspect.

Discussion

The 4D development model by Thiagarajan consists of several stages, including defining, designing, developing, and disseminating. According to (Mulyatiningsih, 2013), the definition stage is also known as the stage of analyzing the needs of students. This stage aims to analyze the needs of students and define the conditions for developing the instrument to be carried out. Meanwhile, at the design stage, the researcher will select the format for the components in the assessment instrument and design the instrument to be developed. At the development stage, the instrument for the assessment of HOTS-oriented mathematics learning outcomes was developed. At the dissemination stage, the instrument had been developed to test the effectiveness of using the HOTS-oriented mathematics learning outcome assessment instrument. Especially for the dissemination stage, the researcher did not do it due to the limited time and costs of the researcher, so the researcher did not reach this stage. Instruments that already have conformity with the theory of instrument preparation are one of the characteristics of a feasible and good instrument to be tested on a limited basis to determine the empirical quality of the instrument (Arikunto, 2009). The developed instruments are then subjected to expert tests/material judges. After carrying out the expert tests, the lecturers as judges will provide suggestions and comments regarding the developed instruments.Instruments are tools used by teachers to collect data when students have gone through the learning process (Hardiani, 2017). With the development of learning assessment instruments that aim to increase students' knowledge and understanding of learning material, the development can create quality questions (Widana, 2017).

The instrument developed and having good validity quality must also distinguish the abilities of students with high and low abilities. Judging from the quality of the discriminatory power of 16 items, the items are in the fairly good category, which means that the items developed are sufficient to distinguish students who have high abilities and low abilities in terms of differentiating power, the category is quite good, meaning that the questions given by smart students can answer. Students who less intelligent some students can answer the question as well. In comparison, the two items are in a good category to distinguish the abilities of smart and less intelligent students, which means that smart students can answer and less intelligent students are unable to answer questions. Questions that can be answered by smart students and students who are not smart enough to be unable to answer mean that the question has distinguishing power and vice versa. The cheat doesn't work (Handayani, 2020). Apart from having discriminatory power, the instrument also needs to have the right level of difficulty. Judging from the level of difficulty, the test has the quality of 1 question in the difficult category, 14 questions in the medium category, and three items in the easy category, this can be interpreted that the difficulty level is in the medium category, which is not too difficult and not too easy. In addition, too easy questions do not stimulate students to enhance their efforts to solve them. On the other hand, too difficult questions will cause students to become desperate, not having the enthusiasm to try again because they are beyond their ability to do the work (Kamarullah, 2017).

In this study, the development of the HOTS-oriented mathematics learning outcome assessment instrument on the Bangun Datar material has a valid, reliable, and appropriate quality to use. Judging from the quality of the content validity of the assessment instrument developed, it obtained a very high category from the results of the expert test. Because the instrument developed based on the assessment of material experts based on the theory of the instrument's preparation already has a match between the indicators and the items, the grammar of the instrument. This is important because content validity determines how important the test is to function as a measuring tool. In line with this, the instrument's content validity relates to how scanned the aspect is designed to measure and the extent to which the items in the instrument reflect the characteristics to be measured (Badriyah et al., 2018). From the trial, data were obtained regarding the students' ability to answer the HOTS instrument. Judging from the quality of the validity of the items, it was found that of the 20 questions, 18 questions were declared valid, and two items were declared invalid. Eighteen items were declared valid because the quality of the test items was good, which could measure according to their function. In comparison, two items were invalid because the quality of the test items was not good which could not carry out its measuring function properly. An assessment instrument that is said to be valid, which means the items have been designed well. In contrast, invalid, which means that the questions designed have not been designed properly in terms of indicators, subject matter, and language (Setemen, 2018).

A good item is seen from the function of the distractor. Judging from the quality of good distractors, at least 5% of the students who took the test were selected. From the results of the calculation of the HOTS-oriented mathematics learning outcome assessment instrument, there were 53 distractors > from 5%, and there were seven distractors < from 5%. This can be interpreted that the quality of the detractor developed is functioning well. A well-functioning distractor means that the distractor can confuse students between the answer key and the distractor. The answer choices are provided with the aim of misleading (deceiving) students who do not understand the material being tested, on the other hand, a distractor that is proven to be ineffective to use, and it is recommended to replace the distractor with a better distractor (Fatimah & Alfath, 2019). After knowing the quality of distractors, it is necessary to know the reliability of the developed instrument. Judging from the results, it shows that the quality of reliability is in the very high category, which means that the questions developed can have consistency, reliability, and stability when tested at different times. Reliability is tested to determine the extent to which a measurement can be trusted because of its constancy as a measuring instrument (Febrianawati, 2018). Based on the results of expert validation, field trials, data analysis, and improvements made, it can be seen that the developed instrument has met the valid and reliable criteria. The quality of the items is good so that the developed instrument can be used as an evaluation tool or an accurate learning assessment on the cognitive aspect.

4. CONCLUSION

Based on the research results, the assessment instrument was declared suitable for use with the quality of the instrument owned based on very high validity and reliability, distinguishing power in the fairly good category, moderate level of difficulty quality, and good distractor quality. The limitation of this study lies in the stages in the 4-D model that does not reach the deployment stage to the teacher if they can use the instrument for assessing higher-order thinking skills in mathematics lessons as an evaluation tool for flat-building material for fourth-grade elementary school. Teachers should also continue to develop higher-order thinking skills instruments on other materials. To the Principal, facilitating teachers to develop instruments to provide school progress and improve learning processes and learning outcomes. For other research, the results of this study should be used as a reference on learning problems and continue to explore more diverse sources as a consideration in developing similar products.

5. REFERENCES

Ahmad, S., Kenedi, A. K., & Masniladevi, M. (2018). Instrumen Hots Matematika Bagi Mahasiswa PGSD. *Jurnal PAJAR (Pendidikan Dan Pengajaran*, *2*(6), 905–912.

Antara, Suwela, I. G. W., Sudarma, I. K., & Dibia, I. K. (2020). The Assessment Instrument of Mathematics Learning Outcomes Based on HOTS Toward Two-Dimensional Geometry Topic. *Indonesian Journal of Educational Research and Review*, *3*(2).

Arikunto. (2009). *Dasar-Dasar Evaluasi Pendidikan Edisi Revisi*. Bumi Aksara.

Astiwi, Tri, K. P., Antara, P. A., & Agustiana, I. G. A. T. (2020). Pengembangan Instrumen Penilaian Kemampuan Berpikir Kritis Siswa SD Pada Mata Pelajaran PPKn. *Jurnal Ilmiah Pendidikan Profesi Guru*, *3*(3).

Badriyah, N. L., Thamrin, A., & Nurhidayati, A. (2018). Analisis Instrumen Penilaian Hasil Belajar Mata Pelajaran Gambar Teknik Siswa Kelas X Sekolah Menengah Kejuruan Program Keahlian Bangunan. *Inndonesian Journal of Civil Engineering Education*, *4*(2).

Budi, S. P. (2014). Pengaruh Pembelajaran Berbasis Masalah Terhadap Kemampuan Berpikir Kritis Siswa Pada Mata Pelajaran Matematika Kelas Iv Gugus Iii Kecamatan Busungbiu. *Journal Mimbar PGSD Universitas Pendidikan Ganesha*. https://doi.org/10.23887/jjpgsd.v2i1.2041.

Budiarta, K., Harahap, M. H., Faisal, & Mailani, E. (2018). *Potret Implementasi Pembelajaran Berbasis High Order Thinking Skills (Hots) Di Sekolah Dasar Kota Medan*.

Budiman, A., & Jailani. (2014). Pengembangan Instrumen Asesmen Higher Order Thinking Skill (Hots) Pada Mata Pelajaran Matematika Smp Kelas Viii Semester 1. *Jurnal Riset Pendidikan Matematika*, *1*, 2,.

Desilva, D., Sakti, I., & Medriati, R. (2020). Pengembangan Instrumen Penilaian Hasil Belajar Fisika Berorientasi HOTS (High Order Thinking Skills) Pada Materi Elastistas dan Hukum Hooke. *Jurnal Kumparan Fisika*, *3*(1), 41–50. https://doi.org/10.33369/jkf.3.1.41-50.

Dharmawati. (2016). Pengembangan Instrumen Asesmen Berpikir Kritis Untuk Siswa Smp Kelas Vii Pada Materi Interaksi Makhluk Hidup Dengan Lingkungan. *Jurnal Pendidikan*, *1*(8).

Farhan, M. (2018). *E-jurnal mitra pendidikan* (Vol. 1, Issue mor 6, Halaman (615-628)).

Fatimah, L. U., & Alfath, K. (2019). *Analisis Kesukaran Soal, Daya Pembeda Dan Fungsi Distraktor*.

Febrianawati, Y. (2018). Uji Validitas Dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jurnal Ilmiah Kependidikan*, *7*(1). https://doi.org/10.18592/tarbiyah.v7i1.2100.

Febrina, F., Bustami, U., & Asnawi, M. (n.d.). Analysis of Reading Comprehension Questions by Using Revised Bloom's Taxonomy on Higher Order Thinking Skill (HOTS. *English Education Journal: Jurnal Pengembangan Pendidikan Dan Pengembangan Pengajaran Bahasa*, *10*(1), 1–15.

Ginting, R., & Haryati, T. (2012). Kepemimpinan dan Konteks Peningkatan Mutu Pendidikan. *Jurnal Ilmiah CIVIS*, *II*(2), 1–17.

Handayani, H. (2015). Pengaruh Pembelajaran Kontekstual Terhadap Kemampuan Pemahaman Dan Representasi Matematis Siswa Sekolah Dasar. *Jurnal Pendidikan Guru Sekolah Dasar*, *1*.

Handayani, S. A. (2020). Humaniora Dan Era Disrupsi Teknologi Dalam Konteks Historis. *Unej E-Proceeding*, *1*(1).

Hanifah, N. (2019). Pengembangan instrumen penilaian Higher Order Thinking Skill (HOTS) di sekolah dasar. *Current Research in Education: Conference Series Journal*, *1*(1).

Hapsari, P., & Kusnindyah, P. (2020). Kemampuan Berpikir Tingkat Tinggi Dalam Pemanfaatan YouTube Sebagai Media Pembelajaran Bahasa Indonesia. In *Jurnal Ilmiah Sekolah Dasar* (Vols. 4, NumberH).

Hardiani, I. N. (2017). Pengembangan Instrumen Penilaian Sikap Sosial Pembelajaran IPS Kelas IV SD. *E-Journal Mitra Pendidikan*, *1*(60), 615–628. https://doi.org/10.1017/CBO9781107415324.004.

Ikhsan, M. K., & Handayani. (2016). the Development of Students ' Worksheet Using Scientific Approach on Curriculum Materials. In *Fourth International Seminar English Language and Teaching (ISELT-4* (pp. 74–87). STKIP Sumatera Barat.

Kamarullah, K. (2017). Pendidikan Matematika di Sekolah Kita. *Al Khawarizmi: Jurnal Pendidikan Dan Pembelajaran Matematika*, *1*(mor 1, Halaman), 21–32.

Kaur Swaran Singh, C., Nor Safinas Raja Harun, R., Arumugam Pillay, L. M., & Singh Masa Singh, T. (2019). The Implementation of Higher Order Thinking Skills for Teaching and Learning Analysis of teachers' implementation of higher-order thinking skills and developing models and modules for teaching writing to weak learners View project Development of a. *The Journal of Social Sciences Research ISSN*, *5*(4), 957–964. https://doi.org/10.32861/jssr.54.957.96.

Kholis, R. A. N. (2017). Analisis Tingkat Kesulitan (difficulty level) soal pada buku sejarah kebudayaan Islam Kurikulum 2013. *Jurnal Pendidikan Agama Islam*, *4*(2).

Koyan. (2012). *kualitatif dalam penelitian. Undiksha Pers*. Alfabeta.

Koyan, I. W. (2011). *Asesmen Dalam Pendidikan*. Universitas Pendidikan Ganesha.

Kristanto. (2020). Pengembangan Soal HOTS (Higher Order Thinking Skills) Terkait Dengan Konteks Pedesaan. *PRISMA, Prosiding Seminar Nasional Matematika*, *3*, 370–376.

Kusainun, N. (2020). Analisis Standar Penilaian Pendidikan di Indonesia. *Jurnal Pendidikan*, *5*(1).

Lestari, R. A., & Subiyanto, Hadisaputro Murbangun, N. (2015). Pembelajaran Berbasis Proyek Dengan Produk Artikel Untuk Meningkatkan Hasil Belajar Siswa. *Chemistry in Education*, *4*(2), 15–21.

Mailani, E. (2019). Order Thinking Skills (HOTS) di Sekolah Dasar Kota Medan. *Jurnal Pembangunan Perkotaan*, *6*(2).

Marzuki, A., & Seri, A. (2017). Development Of Realistic Mathematics Learning Devices to Teach Student Problem Solving Ability Mathematics SMP. *Jurnal: Mosharafa*, *6*, *N0*, 3.

Mislikhah, S. (2020). *Implementasi Higher Order Thinking Skills*.

Mulyatiningsih, E. (2013). *Metode Penelitian Terapan Bidang Pendidikan*. 35,110,114,120,121.

Muslich, M. (2011). *Penilaian Berbasis Kelas dan Kompetensi*. PT Refika Aditama.

Nasrum, A. (2020). Pengembangan Instrumen Evaluasi Pemahaman Konsep Kalkulus Berbasis Komputer. *Jurnal Pendidikan Matematika*, *4*(1).

Niswara, R., Muhajir, M., & Untari, M. F. A. (2019). Pengaruh Model Project Based Learning Terhadap High Order Thinking Skill. *Mimbar PGSD Undiksha*, *7*(: 2).

Nurfillaili, U., T, Y., & Muhammad, A. S. (2016). Pengembangan Instrumen Tes Hasil Belajar Kognitif Mata Pelajaran Fisika Pada Pokok Bahasan Usaha Dan Energi Sma Negeri Khusus Jeneponto Kelas Xi Semester I. *Jurnal Pendidikan Fisika*, *4*(2).

Purnama, Y. I., & Nurdianingsih, F. (2019). The Impact of Higher Order Thinking Skills (HOTS) Instructions in Teaching EFL Speaking Skill from the Perspective of Students' Motivation. *Lingua Cultura*, *13*(4), 313. https://doi.org/10.21512/lc.v13i4.6105.

Putra, H. D. (2017). *Pengembangan Instrumen Untuk Meningkatkan Kemampuan Mathematical Problem*

Posing Siswa Sma.
Rahayu, S., Yusuf, S., & Pranata, O. H. (2020). Pengembangan soal High Order Thinking Skill untuk Meningkatkan Kemampuan Berpikir Tingkat Tinggi Matematika Siswa Sekolah Dasar. *Jurnal Ilmiah Mahasiswa Pendidikan Guru Sekolah Dasar*, *7*(2), 127–137.
Ricardo, R. (2017). Peran Ethnomatematika Dalam Penerapan Pembelajaran Matematika Pada Kurikulum 2013. *LITERASI (Jurnal Ilmu Pendidikan*, *7*(mor 2), 118–125.
Setemen, K. (2018). Pengembangan Dan Pengujian Validitas Butir Instrumen Kecerdasan Logis-Matematis. *Jurnal Pendidikan Teknologi Dan Kejuruan*, *15*(2), 178–187.
Singsh, C., Harun, R., Pillay, L., & Singsh, T. (2019). The Implementation of Higher Order Thinking Skills for Teaching and Learning Analisis. *The Journal of Social Sciences Research ISSN*, *5*(4), 957–964. https://doi.org/10.32861/jssr.54.957.96
Suci. (2020). *Assessment Instruments of Social Attitudes and Social Studies Learning Outcomes for Class V SD on Theme of Lingkungan Sekitar Kita*. https://doi.org/10.23887/jere.v4i3.27486.
Sukla, D., & Dungsungneon, A. P. (2016). Students Perceived Level and Teachers Teaching Strategies of Higher Order Thinking Skills; A Study on Higher Educational Institutions in Thailand. *Journal of Education and Practice*, *7*(12), 211–219.
Trisnawati, W., & Sari, A. K. (2018). Integrasi Keterampilan Abad 21 Dalam Modul Sociolinguistics: Keterampilan 4c (Collaboration, Communication, Critical Thinking, Dan Creativity. *Jurnal Muara Pendidikan*, *4*(2).
Tyas, M. A., Nurkamto, J., Marmanto, S., & Laksani, H. (n.d.). Developing Higher Order Thinking Skills (HOTS) – Based Questions. In *Indonesian EFL Teachers' Challenges* (pp. 52–63). https://doi.org/10.17501/26307413.2019.2106.
Uno, H. B., & Koni, S. (2012). *Assessment Pembelajaran*. PT Bumi Aksara.
Widana, I. W. (2017). Higher Order Thinking Skills Assessment (HOTS. *Jisae: Journal of Indonesian Student*.
Wijayanti, A. (2014). Pengembangan Autentic Assesment Berbasis Proyek Dengan Pendekatan Saintifik Untuk Meningkatkan Keterampilan Berpikir Ilmiah Mahasiswa. *Jurnal Pendidikan IPA Indonesia*, *3*(2), 102–108.
Wirayasa, I. D. G. P., Darmayasa, I. P., & Satyawan, I. M. (2020). Pengembangan Instrumen Penilaian Hasil Belajar Ranah Kognitif Model 4D Pada Materi Sepak Bola Berdasarkan Kurikulum 2013. *Jurnal Pendidikan Jasmani Olahraga Dan Kesehatan*, *8*(3).
Yuliandini, N., Hamdu, G., & Respati, R. (2019). Pengembangan Soal Tes Berbasis Higher Order Thinking Skill (HOTS. *Taksonomi Bloom Revisi Di Sekolah Dasar. Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, *6*(1).
Zainal, F. M. (2018). Strategi Pengembangan Soal Higher Order Thinking Skill (Hots) Dalam Kurikulum 2013. *Journal Of Islamic Religious Education. Vol.II*, *1*.