Jurnal Ilmiah Sekolah Dasar
Volume 5, Number 2, 2021 pp. 231-241
P-ISSN: 2579-3276 E-ISSN : 2549-6174
Open Access: ttps://ejournal.undiksha.ac.id/index.php/JISD/index

HOTS-based Instrument for Assessing Students Science Learning Outcomes in Elementary School

Ni Putu Sri Diah Anggraeni[1]*, Gede Wira Bayu[2] , I Gde Wawan Sudatha[3]

[1,2,3] *Elementary School Teacher Education Study Program, Universitas Pendidikan Ganesha, Singaraja, Indonesia*

ARTICLE INFO

Article history:
Received March 08, 2021
Revised March 11, 2021
Accepted April 30, 2021
Available online May 25, 2021

Kata Kunci:
Instrumen Penilaian IPA , HOTS, Model 4-D.

Keywords:
Assessment Instrument, Science, 4-D Models

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan instrumen penilaian tes hasil belajar IPA berbasis HOTS (Higher Order Thinking Skill). Penelitian ini merupakan penelitian pengembangan yang menggunakan model 4D (four-D models) dengan tahap pengembangan yang terdiri dari define, design, develop, dan desseminate. Penelitian pengembangan ini hanya dilakukan sampai tahap develop. Subjek penelitian ini adalah instrumen penilaian hasil belajar IPA berbasis HOTS berupa kisi-kisi, dan lembar tes pilihan ganda. Data yang diperoleh menggunakan metode wawancara, observasi, dan tes. Validasi instrumen penilaian dilakukan oleh dua orang ahli materi menggunakan lembar validasi dan 78 siswa untuk uji coba terbatas menggunakan instrumen tes objektif pilihan ganda. Hasil analisis intrumen penilaian hasil belajar IPA berbasis HOTS memiliki validitas sebesar 0,90 yang berada pada kategori sangat tinggi, reliabilitas sebesar 0,81 yang berada pada kategori sangat tinggi. Analisis daya beda mendapatkan hasil 2 butir soal dengan kriteria sangat baik, 14 butir soal dengan kriteria baik, dan 9 butir soal dengan kriteria cukup. Pada uji tingkat kesukaran hasilnya sebanyak 12 soal berada pada kategori mudah, dan 13 soal berada pada kategori sedang. Analisis uji kualitas pengecoh mendapatkan hasil 63 pengecoh berada pada taraf >5% yang artinya pengecoh berfungsi dengan baik dan 12 pengecoh berada pada taraf ≤5% yang artinya pengecoh tidak berfungsi dengan baik. Hasil tersebut menunjukkan instrument penilaian tes hasil belajar IPA berbasis HOTS yang dikembangkan valid dan reliabel serta layak digunakan sebagai instrumen penilaian pada materi macam-macam gaya.

ABSTRACT

This study aimed to develop an instrument for assessing science learning outcomes based on HOTS (Higher Order Thinking Skills). This research was development research that uses a 4D model (four-D model) with development stages consisting of define, design, develop, and distribute. This development research was only carried out until the development stage. The subject of this research was the HOTS-based science learning outcome assessment instrument in the form of grids and multiple-choice test sheets. Data were obtained by using interviews, observation, and test methods. The validation of the assessment instrument was carried out by two material experts using a validation sheet and 78 students for a limited trial using multiple-choice objective test instruments. The results obtained will be analyzed for validity, reliability, discriminatory power, level of difficulty, and quality of distractors. The results of the analysis of the HOTS-based science learning outcome assessment instrument had a validity of 0.90 which was in the very high category, reliability of 0.81 which was in the very high category. Distinguishing analysis obtained 2 items with very good criteria, 14 items with good criteria, and 9 items with sufficient criteria. In the difficulty level test, 12 questions were obtained in the easy category and 13 questions in the medium category. Analysis of the distractor quality test found 63 distractors were at the >5% level, which means the distractors are functioning well and 12 distractors were at the 5% level, which means the distractors were not functioning properly. These results indicated that the HOTS-based science learning outcome test assessment instrument developed was valid and reliable and is suitable for use as an assessment instrument in various styles of material.

Corresponding author
*E-mail addresses: : sridiahanggraeni@gmail.com

1. INTRODUCTION

Assessment is an important component in the learning process. In essence, assessment is a component that is used to obtain information about children's learning processes that are used to determine the effectiveness of learning, so that assessment is said to be a very important component in the learning process (Ahmad, 2017; Kusaeri et al., 2017; Zhang, 2020). Assessment is all activities that include collecting and processing information about decision-making on learning outcomes (Black & Wiliam, 2018; Kusainun, 2020; Leong et al., 2018; Salamah, 2018). In addition, assessment is a process carried out to measure the implementation of the learning process (Widiyanto & Istiqomah, 2020). Assessment is an important thing because it has a very important influence on learning including improving the learning process between teachers and students and also make learning take place effectively and can identify learning needs which can later be used to help improve the continuous learning process (Anandan, 2015). In the implementation of the assessment must adjust between what will be measured and using the right measuring instrument. One of the measuring tools that can be used to conduct an assessment is an instrument. The instrument is said to be a tool used to collect data on learning outcomes and evaluation of the learning process that is used to facilitate a task or job to achieve a goal effectively (Desilva et al., 2020; Gaol et al., 2017; Hardiani, 2017). In addition, the instrument is also said to be an assessment tool that can be used to assess the learning achievement of students, instruments that are appropriate to be used to conduct an assessment are instruments that can determine the extent to which these students achieve the learning objectives that have been set and pay attention to the level of students' thinking abilities (Wahida, 2018). The principle in making the instrument is that the instrument made must be objectively based on clear, systematic, continuous, integrated, open, and fair procedures (Kholis, 2017; Umami, 2018).

What happened based on the results of interviews with fourth-grade teachers at Gugus Lompa Batang Elementary School, Melaya District on November 6, 2020, most of the teachers stated that in the preparation of assessment instruments in the cognitive domain such as the end of semester assessment the teacher always tried to adjust to the cognitive level of students, but in making assessment instruments on higher cognitive teachers are still having difficulties and constrained by time. Some teachers also stated that due to time constraints the teacher only made the questions, without being equipped with a grid. In addition, based on the analysis of document studies on the assessment of daily tests and midterm tests of science content for class IV for the 2020/2021 academic year at SD Gugus Lompa Batang it was found that most of the assessment instruments made could only measure students' cognitive abilities at cognitive levels C1 to C3, while the cognitive level C4 to C6 has not been achieved optimally. Teachers often find time constraints in the preparation of instruments, causing a mismatch between what is measured and the measuring instrument used (Mahmuda et al., 2017). Another problem that was found was the imbalance between ideal conditions and real conditions such as the less complex cognitive level contained in the compiled questions (Antara et al., 2020). Most of the question grids prepared by the teacher between KD and indicators are not appropriate and the validity of the items has not been tested so that the instrument prepared is not yet clear on the level of feasibility as an assessment tool (Sutami et al., 2021).

Based on the problems found, if not followed up, it will have a negative impact on the assessment process in the learning process. One solution that can be done is the development of appropriate and quality assessment instruments. The alternative is in line with research that has been done previously, namely the development of an instrument for learning mathematics outcomes that obtain valid and reliable instrument results so that it can be used to measure students' mathematics learning outcomes. (Pratiwi & Mahfud, 2020; Suarsih et al., 2020). Furthermore, the development of cognitive assessment instruments that obtain results that are very valid, reliable, and have practical value (Hamid, 2016; Nurfillaili et al., 2016). Development of motivational assessment instruments and science learning outcomes that produce valid and reliable instruments (Mudanta et al., 2020). The development of the instrument was carried out, namely the development of an evaluation instrument for understanding computer-based calculus concepts that obtained valid and practical results that could be used to measure students' ability to understand calculus concepts. Furthermore, the development of critical thinking ability assessment instruments in Civics subjects obtained valid and reliable results so that the instrument is suitable for use in conducting assessments (Astiwi et al., 2020).

Based on the problems and solutions that have been done, a solution is needed that can overcome the problems or findings and shortcomings of the previous solutions. So it is necessary to develop the HOTS-based science learning outcomes instrument for elementary school students on various styles of material. *Higher-Order Thinking Skills (HOTS)* or higher-order thinking skills are skills to connect ideas and facts, analyze, explain, hypothesize, synthesize or arrive at the stage of concluding to solve problems (Lestari, 2016). *Higher Order Thinking Skills (HOTS)* or higher-order thinking skills are said to be very in line with the demands of the 2013 curriculum, namely students can not only know, understand and apply

but students must also be able to analyze, evaluate, and even create in learning (Yuliandini et al., 2019). The development of the HOTS-based learning outcome assessment instrument is very important to be realized to obtain a suitable instrument to use in assessing a good cognitive domain. The assessment instrument developed has a novelty that can distinguish it from previous research, namely differences in the material developed, namely various styles, the development of this instrument also places more emphasis on higher cognitive levels, namely from C3 to C6. The advantages of the developed instrument are that it can measure students' thinking (cognitive) abilities at a high level, the material selected on the instrument has been adjusted to the student's daily activities, and is in accordance with the demands of the curriculum, namely the 2013 curriculum which requires students to be able to think at a higher level. The development of an assessment instrument in the cognitive domain aims to develop an assessment instrument for science learning outcomes based on *HOTS* on various styles of material in grade IV of the Gugus Lompa Batang Elementary School. The development of this instrument is expected to assist teachers in conducting quality assessments using tested and appropriate assessment instruments.

2. METHOD

This type of research is developmental research. This research is designed to assist teachers in choosing and developing an appropriate and appropriate instrument to measure students' abilities in the science learning process. The development model used in 4D. Thiagarajan stated that the 4D model consists of define, design, develop, and disseminate stages (Dewi & Akhlis, 2016). However, the dissemination stage cannot be carried out due to time and financial constraints. The implementation of the 4D model in this study consists of (1) the define stage, which is the initial stage or becomes the basis for conducting research, at this stage the identification of needs and field information collection related to the instrument to be developed is carried out. (2) The design stage is the stage for planning research. At this stage, it is designed to develop an initial framework of the instrument to be made, namely the *HOTS (Higher Order Thinking Skill)*-based science learning outcome test instrument for grade IV students. (3) The development stage is the stage of producing an instrument for assessing learning outcomes based on HOTS class IV which has passed the guidance and revision process through expert trials/judges and field trials. The subjects in this study were the *HOTS*-based science learning outcomes test assessment instrument on theme 7 the beauty of diversity in my country on the topic of various styles of fourth-grade elementary schools in the Lompa Batang cluster, Melaya District.

Data collection methods used include interviews, observation, and tests. The interview is a method that contains a set of structured questions that are given to informants to dig up information (Adhimah, 2020). Observation is a technique used to collect data by recording an event with the help of an instrument to achieve a research objective (Syamsudin, 2014). The purpose of the interview and observation method is to obtain initial information to develop the product. A test is a tool used to find out or measure something with predetermined stages and rules (Suharman, 2018). This test method aims to obtain data from field test results by administering tests to 78 students in the form of multiple-choice objective tests. This development uses data analysis methods including analysis of validity, reliability, discriminating power, level of difficulty, and quality of instrument distractors. Validity analysis consists of two analyzes, namely content validity analysis and item validity analysis. Content validity analysis using the Gregory formula of 2x2 cross-tabulation through expert judgments according to their expertise. In this development, experts come from lecturers in the Faculty of Education, the Ganesha University of Education who have expertise in the field of science. The content validity values obtained are then presented based on the content validity coefficient categories in Table 1.

Tabel 1. Category Content Validity Coefficient

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

(Arikunto, 2016)

Item validity analysis is the analysis used to interpret the number of valid or not test items being tested on students. In this development, multiple-choice tests are used. The multiple-choice test items are called dichotomous tests because the multiple-choice test scores are in the form of a dichotomous scale,

namely 1 (one) and 0 (zero). A score of 1 is given for the correct answer to the item, while 0 is given for the wrong answer. The item validity analysis used the point-biserial correlation (Ypbi) technique. The test items can be declared valid if the r count is greater than the r table with a significance level of 5%. The reliability test was carried out on valid items only. reliability is a level of consistency or stability of the results on a measurement (Zahra & Nofha, 2018). Analysis of the reliability test using the KR-20 formula. The reliability coefficients obtained are then compared with the criteria in Table 2.

Table 2. Criteria of Instrument Reliability

Coefficient Interval	Criteria
0,81 – 1,00	Very high
0,61 – 0,80	High
0,41 – 0,60	Medium
0,21 – 0,40	Low
0,00 – 0,20	Very low

(Koyan, 2011)

Differential power test is done after doing reliability test. Distinguishing power is a measure that shows the level of ability of the items that distinguish the high-achieving group and the low-achieving group within the scope of the test takers (Fatimah & Khairudin, 2019). The discriminatory power test was only carried out on test items that were declared valid (25 items). To determine the upper and lower groups of the total test-takers, using 27% of the total test-takers (sample) with the highest score as the upper group, and 27% of the total test-takers (sample) with the lowest score as the lower group. The results of the calculation of the different power obtained are then compared with the criteria in Table 3.

Table 3. Criteria of Different Power

Coefficient Interval	Criteria
0,71 – 1,00	Very good
0,40 – 0,70	Good
0,20 – 0,39	Pretty good
0,00 – 0,19	Not good

(Koyan, 2011)

The difficulty level test is carried out on valid items only. The level of difficulty is a level of easy or difficult items for a group of students (Pratiwi & Mahfud, 2020). The difficulty level of the test is also expressed as a number that shows the average proportion of the test that can answer correctly. The criteria for the level of difficulty can be seen in Table 4.

Table 4. Criteria of level difficulty

Coefficient Interval	Criteria
0,00 – 0,29	Hard
0,30 – 0,70	Medium
0,71 – 1,00	Easy

(Candiasa, 2010)

The distractor quality is used to determine whether the distractor has functioned as a distractor well or not. the distractor is functioning well if more than 5% of the test takers have been selected (p > 5%) and if it is less or equal to 5% (p 5%) it means that the distractor is not functioning properly (Uno & Koni, 2012). Determining the quality of the distractor uses four alternative answers so that three of them function as distractors. The quality criteria for distractors can be seen in Table 5.

Table 5. Distractor Quality Test Results Analysis

Distractor	Criteria
(p > 5%)	Works fine
(p ≤ 5%)	Not functioning properly

3. RESULT AND DISCUSSION

The defined stage is the initial stage or becomes the basis for conducting research. At this stage, information was obtained through interviews, most of the teachers stated that in the preparation of assessment instruments in the cognitive domain such as at the end of the semester assessment the teacher always tried to adjust to the cognitive level of students, but in making assessment instruments on higher cognitive the teacher still had difficulties and time constraints. Some teachers also stated that due to time constraints the teacher only made the questions, without being equipped with a grid. In addition, based on observations and document studies, it was found that most of the questions compiled were only at the low cognitive level from C1 to C3 but had not yet reached the high-level cognitive level, namely from C4 to C6. This is certainly not in line with the demands of the 2013 curriculum which requires students to be able to think at a high cognitive level. The design stage is the stage for planning research. At this stage, it is designed to develop an initial framework of the instrument to be made. At this stage, the instruments are arranged based on basic competencies (KD) in science subjects, especially theme 7, the beauty of diversity in my country. Basic competence is limited to KD 3.3 identifying various styles which are developed into 13 question indicators. The next stage is the preparation of the HOTS-based science assessment instrument grid, which is presented in Table 6.

Table 6. Blueprint of *HOTS* Based Science Learning Outcomes Assessment Instruments

No.	Basic competencies	Indicator	Cognitive Level	Number of questions	number of questions
1.	3.3 Identify various types of forces, including muscle force, electric force, magnetic force, gravitational force, and frictional force.	3.3.1 Applying surfaces of objects that are easy and difficult to hold	C4	1 2	2
		3.3.2 Analyzing the causes of moving objects in everyday life	C4	3 4	2
		3.3.3 Evaluate the occurrence of muscle force events in daily life	C5	5	1
		3.3.4 Analyzing examples of muscle force events in everyday life	C4	6 7 8	3
		3.3.5 Combining tools used to generate dynamic and static electricity every single day	C6	9 10	2
		3.3.6 Analyzing static and dynamic electric forces in everyday life	C4	11 12	2
		3.3.7 Analyzing the working principle of dynamic electric current in electronic objects	C4	13 14 15	3
		3.3.8 Evaluating electronic devices and their functions.	C5	16 17	2
		3.3.9 Analyzing magnetic force events in everyday life	C4	18 19	2
		3.3.10 Analyze the benefits of gravity in everyday life.	C4	20 21	3

No.	Basic competencies	Indicator	Cognitive Level	Number of questions	number of questions
				22	
		3.3.11 Analyze the advantages and disadvantages of frictional events in everyday life.	C4	23 24	2
		3.3.12 Analyze the contents of two magnets if they are brought close to the same or different poles	C4	25	1
		3.3.13 Analyze objects that can be attracted by magnets	C4	26 27	2
		AMOUNT		**27 item**	

The next stage is the initial design stage. The initial design is the stage carried out to develop the initial framework in the preparation of the instrument, at this stage the preparation of the HOTS-based science learning outcome assessment instrument in the form of a multiple-choice test with 27 questions. Here are 3 examples of questions from all the questions presented in Figure 1.

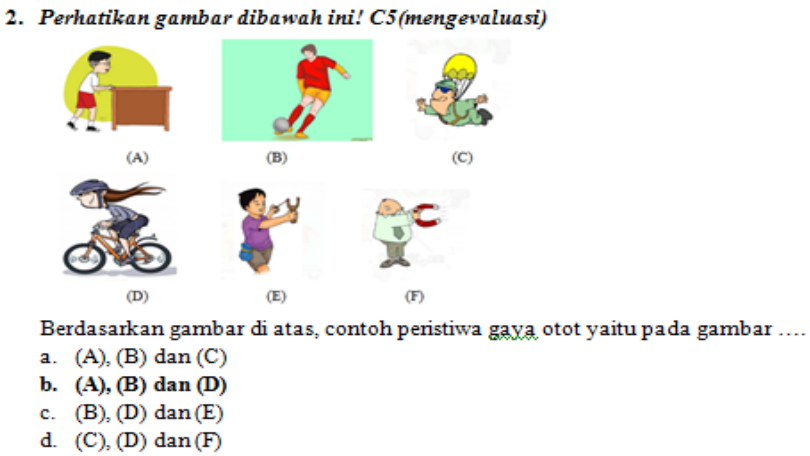

Figure 1. The example of the *HOTS* question

The development stage is the stage of producing an instrument for assessing learning outcomes based on HOTS class IV which has gone through the process of guidance and revision through expert trials/judges and field trials. Expert trials/judges were carried out using 2 science expert lecturers within the Undiksha Faculty of Education. The results of the assessment of the material expert trials/judges are presented in table 7.

Table 7. Expert Judge

Expert/*Judges* 1		Expert/*Judges* 2	
Relevant	**Irrelevant**	**Relevant**	**Irrelevant**
1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30,	18,19	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 22, 23, 24, 25, 26, 27, 29, 30,	18,19,28

Based on the expert test results/judges obtained, then the results are analyzed using the 2x2 cross-tabulation Gregory formula. Based on the analysis, the results of content validity were 0.90 compared to the content validity coefficient category which was in the very high category. The results obtained show that the *HOTS*-based science learning outcome assessment instrument has very high validity. After the material experts/judges fill out the validation sheet, the experts/judges provide comments and suggestions on the assessment instruments that have been prepared. Comments and suggestions are presented in Table 8.

Table 8. IPA Expert Comments and Suggestions

Expert I	Expert II
Question number 1 does not match C4 cognitive realm, this is not included about analyzing. Question number 8, pay attention to options b replace the word wardrobe with a wardrobe. Question number 18 is not by C5 (analyze) Question number 19 is irrelevant because it hasn't Corresponding with questions C5 (analyze).	Questions 1 and 2 do not characterize analytical questions, replace options with pictures. Problem number 3, the word above in the question sentence is replaced with the word picture A. Pushing a table and a car is an example of the effect of STYLE. Question number 5, make an analysis problem by sorting the right pictures! Initial options must use lowercase letters, except for Person Names, Places, and so on. Question number 6, A, and D are correct answers Question number 9, dynamic/static electricity? match the indicators. Questions number 18 and 19, Irrelevant because of the repetition of words Question number 28 should not make the question have the opposite word, because it is a connection from the previous question! Question number 29 for questions like number 25, 26! Question number 30 is for distractors such *except.*

After the expert validates, the next step is to test the instrument. The limited trial is conducted by testing the HOTS-based science learning outcome test assessment instrument for fifth-grade students one level higher. In this study, the instrument is tested in the form of a *HOTS*-based learning outcome test to four different elementary schools, namely SD Negeri 2 Manistutu with a total of 16 students, SD Negeri 3 Manistutu with a total of 22 students, SD Negeri 4 Manistutu with a total of 24 students, and SD Negeri 6 Manistutu with a total of 16 students, the total number of students are seventy-eight. The limited trial is carried out to determine the feasibility of the learning outcomes test instrument that has been tested. The data from the test results are analyzed for item validity, discriminating power, level of difficulty, and distracting quality. The test results have analyzed the validity of the items analyzed using the biserial point correlation coefficient formula assisted by Microsoft Office Excel 2007 presented in Table 9.

Table 9. Results of Analysis of the Validity of Science Learning Outcomes

Item Validity	Number of Questions	amount	Percentage
Valid	2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26	25	93%
Invalid	**4, 27**	**2**	**7%**

Based on table 9 above, the results of the validity of the HOTS-based science learning outcomes instrument of 27 items, as many as 25 items with a percentage of 93% are valid and 2 items with a percentage of 7% are invalid. Items that are declared invalid are not used as instruments.

Reliability analysis is used to determine the level of consistency of the instrument's answers. A good test instrument accurately has consistent answers. Reliability testing is carried out using items that have been declared "valid" and questions that have "invalid" criteria are not analyzed for reliability. This learning outcome score is a dichotomous score calculated using the KR-20 formula. Using the help of the Microsoft Office Excel 2007 application. The instrument reliability results obtained are 0.81. The results are converted into reliability criteria according to Koyan (Koyan, 2011) is very high reliability. The discriminatory power test is only carried out on test items that are declared valid (25 items). To determine the upper and lower groups of the total test-takers, using 27% of the total test-takers (sample) with the highest score as the upper group, and 27% of the total test-takers (sample) with the lowest score as the lower group. The discriminatory power test is calculated using the help of the Microsoft Excel 2007 application. The results of the calculation of the differentiating power of the HOTS-based science learning outcome test can be seen in Table 10.

Table 10. Different Power Test Results

Different Power Criteria	Number of questions	amount
Very good	3, dan 13	2
Good	2, 5, 6, 7, 9, 14, 15, 16, 18, 20, 21, 24, 25, dan 26	14
Pretty good	1, 8, 10, 11, 12, 17, 19, 22, dan 23	9

Based on table 10 above, the results of the analysis of the differentiating power of the *HOTS*-based science learning outcomes instruments are converted into differentiating power criteria according to Koyan (Koyan, 2011) as many as 2 items with very good criteria, 14 items on good criteria, and 9 items on very good criteria. The analysis of the item difficulty level test is only carried out on items that were declared valid (25 items). The difficulty level test is calculated using the help of Microsoft Office Excel 2007. The results of the calculation of the difficulty level of the HOTS-based science learning outcome test can be seen in Table 11.

Table 11. Test Results Level of difficulty

Difficulty Category	Number of Questions	Amount	Percentage
Easy	1, 2, 3, 9, 10, 17, 18, 19, 20, 23, 24, 25	12	48%
Medium	5, 6, 7, 8, 11, 12, 13, 14, 15, 16, 21, 22, 26	13	52%

Based on table 11 above, the results of the analysis of the difficulty level of the *HOTS*-based science learning outcomes instrument are converted into the criteria for the level of difficulty according to Candiasa (Candiasa, 2010) as many as 12 items on easy criteria, and 13 items on medium criteria. The next analysis carried out is a distractor quality analysis. The distractor quality analysis is conducted to determine if the distractor is functioning properly if more than 5% of the test takers had been selected (p > 5%) and if it was less or equal to 5% (p 5%) it meant that the distractor was not functioning properly. Determining the quality of the distractor using four alternative answers so that three of them function as distractors. The quality analysis of the distractor is carried out with the help of the Microsoft Office Excel 2007 application, the results of the analysis of the quality of the distractors of the *HOTS*-based science learning instrument as many as 63 distractors function well, and 12 distractors do not function properly. The dissemination stage, at this stage, cannot be carried out due to time and financial constraints.

Based on the results obtained, the *HOTS*-based science learning outcomes instrument has been tested for validity, reliability, discriminating power, level of difficulty, and quality of distractors. So, it can be said that the *HOTS*-based science learning outcomes instrument can be applied because it already meets the requirements of a good assessment instrument used in conducting assessments. Instruments are often used in the process of collecting data, this is in line with previous researchers using instruments to collect data on mathematics learning outcomes (Wahida, 2018). In addition, the instrument is also used to collect data on mathematics learning outcomes and learning anxiety for fourth-grade elementary school students (Suarsih et al., 2020). Instruments are needed so far in the learning process because most of the results of the needs analysis that have been carried out in the learning process appear to use instruments. If the instrument used is not good or not by what will be measured, it will affect the assessment process. Based on the findings in the field, most of the assessment instruments produced are not by the existing assessment principles. The general principles that must be met in the assessment are valid, educational,

sustainable, meaningful, comprehensive, and competency-oriented (Muslich, 2011). Assessment must also be objectively based on clear, systematic, continuous, integrated, open, and fair procedures (Kholis, 2017; Umami, 2018). If the assessment does not apply these principles and is left alone, it will affect learning, because assessment is a benchmark in the learning process.

This development was designed as an instrument for assessing science learning outcomes based on *HOTS* on various styles of material using multiple-choice tests. The material of various styles was chosen in this development because in this material the HOTS-based multiple-choice instrument was designed. After all, the multiple-choice test is very suitable to be used to conduct assessments with a very large number of participants or a mass nature. In addition, the multiple-choice test is objective, covers a broad scope of material, and has a high level of validity and reliability because it is easy to know in terms of the level of knowledge that has been mastered or not yet mastered by students. The multiple-choice test in this development is based on *HOTS* because it is adapted to the demands of the 2013 curriculum which requires students to think at a high cognitive level, namely at cognitive levels C4 to C6.

The results of developing *HOTS*-based science learning outcomes instruments get valid and reliable results. It is declared valid and reliable because the instrument can be proven, valid, and based on clear and consistent procedures if repeated measurements are made by the principles of assessment. The results of this study are supported by previous research, namely the development of learning outcomes assessment instruments on soccer material which get very high validity and high-reliability results, different power tests, and levels of difficulty which get varying data test results (Wirayasa et al., 2020). Furthermore, research shows that the instrument developed is valid and reliable based on expert tests and validity tests with empirical data and with a Kuder Richardson reliability coefficient (KR-20) of 0.84 (Agustika, 2018). Developing an instrument for cognitive physics learning outcomes that obtained 62 valid questions, has a reliability value of 0.93 which is in the very high category so that the developed instrument is feasible to use for assessing (Ihwan et al., 2019).

The learning outcome assessment instrument is indeed feasible to be developed because the assessment instrument is used to obtain the desired information. This is in line with the opinion of previous researchers who stated that an instrument is a measuring tool used to assess something in the context of collecting data in obtaining the desired information (Paulina & Ertikanto, 2014). So that, an educator needs an assessment instrument to obtain information about student development in learning. A good instrument is an instrument that has validity, reliability, and practicality values (Aji & Winarno, 2016). In research on the development of *HOTS*-based science learning outcomes instruments, the criteria for validity and reliability are very high. Based on previous research, this research has up-to-date material in various styles and the advantages of the instruments developed are being able to measure students' (cognitive) thinking skills at a high level, the material selected on the instrument has been adapted to students' daily activities. and is by the demands of the curriculum, namely the 2013 curriculum which requires students to be able to think at a higher level. In addition, this development in the analysis stage is more complex which consists of the analysis of validity, reliability, discriminating power, level of difficulty, and quality of distractors compared to previous research which was only limited to analysis of validity and reliability. The results of this development have implications for the availability of appropriate science content cognitive domain assessment instruments because they have been tested for quality. Quality consisting of validity, reliability, discriminating power, level of difficulty, and quality of distractors so that it is feasible to use in measuring or assessing in the learning process.

4. CONCLUSION

Based on the results of research and discussion, it can be stated that the instrument of science learning outcomes has been tested for validity and reliability which is in the very high category. In addition, this instrument has been tested for different power levels, difficulty levels, and distracting qualities. Through the results obtained, the *HOTS*-based science learning outcome assessment instrument can be said to be feasible and accurate to use to assess science learning outcomes on various styles of material.

5. REFERENCES

Adhimah, S. (2020). Peran Orang Tua Dalam Menghilangkan Rasa Canggung Anak Usia Dini (Studi Kasus di Desa Karangbong RT. 06 RW.02 Gedangan-Sidoarjo. *Jurnal Pendidikan Anak*, *9*(1).

Agustika, G. N. S. (2018). Pengembangan Konstruksi dan Validasi Tes Konsep Dasar Matematika. *Journal of Education Technology*, *2*(1). https://doi.org/10.23887/jet.v2i1.13805.

Ahmad, K. (2017). Penilaian Pembelajaran Tematik di Madrasah. *Jurnal Pedagogik*, *4*(2).

Aji, B. S., & Winarno, M. E. (2016). Pengembangan Instrumen Penilaian Pengetahuan Mata Pelajaran Pendidikan Jasmani Olahraga dan Kesehatan (PJOK) Kelas VIII Semester Gasal. *Jurnal Pendidikan*, *1*(7), 1449–1463.

Anandan, K. (2015). *Assesment for Learning in Technology Enhaced Learning Research Themes*. Bharathidasan University. https://doi.org/10.1007/978-3-319-02600-8_12.

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, S. (2016). *Prosedur Penelitian Suatu Pendekatan Praktik*. Rineka Cipta.

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).

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy and Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807.

Candiasa, I. M. (2010). Pengujian instrumen penelitian disertai aplikasi ITEMAN dan BIGSTEPS. *Singaraja: Unit Penerbitan Universitas Pendidikan Ganesha*.

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.

Dewi, N. R., & Akhlis, I. (2016). Pengembangan perangkat pembelajaran IPA berbasis pendidikan multikultural menggunakan permainan untuk mengembangkan karakter siswa. *USEJ - Unnes Science Education Journal*, *5*(1).

Fatimah, L. U., & Khairudin, A. (2019). Analisis Kesukaran Soal, Daya Pembeda dan Fungsi Distraktor. *Jurnal Komunikasi Dan Pendidikan Islam*, *8*(2).

Gaol, P. L., Khumaedi, M., & Masrukan. (2017). Pengembangan Instrumen Penilaian Karakter Percaya Diri pada Mata Pelajaran Matematika Sekolah Menengah Pertama. *Jurnal of Educational Research and Evaluation*, *6*(1). https://doi.org/10.1016/j.tate.2020.103193.

Hamid, M. A. (2016). Pengembangan Instrumen Penilaian Hasil Belajar Siswa Berbasis TIK pada Pembelajaran dasar Listrik Elektronika. *Jurnal Ilmiah Pendidikan Teknik Elektro*, *1*(1).

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.

Ihwan, M. Al, Sari, S. S., & Ali, M. sidin. (2019). Pengembangan Instrumen Tes Hasil Belajar Kognitif Fisika Kelas XI MIA SMA Negeri 5 Pinrang. *Jurnal Sains Dan Pendidikan Fisika*, *15*(2).

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, I. W. (2011). *Asesmen Dalam Pendidikan* (Undiksha Press (Ed.)).

Kusaeri, Mutaqin, Z., & Mochamad. (2017). Pengembangan Instrumen Penilaian Tes Tertulis Bentuk Uraian Untuk Pembelajaran PAI Berbasis Masalah Materi FIQH. *Jurnal Pemikiran Dan Penelitian Pendidikan*, *15*(1).

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

Leong, W. S., Ismail, H., Costa, J. S., & Tan, H. B. (2018). Studies in Educational Evaluation Assessment for learning research in East Asian countries. *Studies in Educational Evaluation*, *59*(September), 270–277. https://doi.org/10.1016/j.stueduc.2018.09.005.

Lestari, A. (2016). Pengembangan Soal Tes Berbasis Hots Pada Model Pembelajaran Latihan Penelitian Di Sekolah Dasar. *Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, *3*(1).

Mahmuda, A., Kartika, I., & Raden, O. (2017). Pengembangan dan Uji Coba Instrumen Penilaian Hasil Belajar IPA SMP/MTs Kelas VII Pada Materi Karakteristik Zat. *Jurnal Berkala Fisika Indonesia*, *9*(1).

Maulida, I., Dibia, I. K., & Astawan, I. G. (2020). The Development of Social Attitude Assesment Instrument and Social Studies Learning Outcomes Grade IV on Theme of Indahnya Keragaman di Negeriku. *Indonesian Journal of Educational Research and Review*, *3*(2).

Mudanta, K. A., Astawan, G., & Labajayanta, N. (2020). Pengembangan Instrumen Penilaian Motivasi Belajar dan Hasil Belajar IPA Siswa Kelas V SDN 1 Sepang Kelod Kecamatan Busungbiu Tahun Pelajaran 2019/2020. *Jurnal Mimbar Ilmu*, *25*(2). https://doi.org/10.23887/mi.v25i2.26611.

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).

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).

Paulina, R., & Ertikanto, C. (2014). *Pengembangan Instrumen Penilaian Pembelajaran Sains Bermuatan Nilai Ketuhanan dan Kecintaan Terhadap Lingkungan*. Jurnal Pembelajaran Fisika Universitas Lampung.

Pratiwi, A. A., & Mahfud, E. (2020). Pengembangan Instrumen Evaluasi Pembelajaran Matematika Tipe PISA Berkarakteristik Kebudayaan Lokal. *Jurnal Pendidikan Rafa*, *6*(1).

Salamah, U. (2018). Penjaminan Mutu Penilaian Pendidikan. *Jurnal Evaluasi*, *2*(1).

Suarsih, A., Arnyana, I. B. P., & Made, A. (2020). Pengembangan Instrumen Hasil Belajar Matematika dan Kecemasan Belajar Siswa Kelas IV Gugus III Abiansemal Tahun Ajaran 2019/2020. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, *10*(1).

Suharman. (2018). Tes Sebagai Alat Ukur Akademik. *Jurnal Ilmiah Pendidikan Agama Islam*, *10*(1).

Sutami, N. K. A., Dantes, N., & Arnyana, I. B. P. (2021). Pengembangan Instrumen Hasil Belajar IPA dan Kemampuan Metakognitif Siswa Kelas V SD. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, *11*(1).

Syamsudin, A. (2014). Pengembangan Instrumen Evaluasi Non Tes (Informal) untuk Menjaring Data Kualitatif Perkembangan Anak Usia Dini. *Jurnal Pendidikan Anak*, *3*(1).

Umami, M. (2018). Penilaian Autentik Pembelajaran Pendidikan Agama Islam dan Budi Pekerti dalam Kurikulum 2013. *Jurnal Kependidikan*, *6*(2).

Uno, H. B., & Koni, S. (2012). *Assessment Pembelajaran*. PT Bumi Aksara.

Wahida, I. N. (2018). Pengembangan instrumen penilaian hasil belajar Matematika Siswa berdasarkan teori al Mawardi. *Jurnal Pendidikan Matematika*, *3*(1).

Widiyanto, D., & Istiqomah, A. (2020). Evaluasi Penilaian Proses dan Hasil Belajar Mata Pelajaran PPKn. *Citizenship Jurnal Pancasila Dan Kewarganegaraan*, *8*(1).

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).

Zahra, R. K., & Nofha, R. (2018). Pengaruh Celebrity Endorser Hamidah Rachmayanti Terhadap Keputusan Pembelian Produk Online Shop Mayoufit di Kota Bandung. *Jurnal Lontar*, *6*(1).

Zhang, X. (2020). Assessment for learning in constrained contexts: How does the teacher's self-directed development play out? *Studies in Educational Evaluation*, *66*(November 2019), 100909. https://doi.org/10.1016/j.stueduc.2020.100909.