JPSD Vol. 6 No. 2, September 2020
ISSN 2540-9093
E-ISSN 2503-0558

DEVELOPMENT OF ASSESSMENT INSTRUMENTS BASED ON HIGHER ORDER THINKING SKILLS (HOTS) FOR ELEMENTARY SCHOOL STUDENTS

Suratmi, Laihat, Asnimar

Sriwijaya University

suratmi@fkip.unsri.ac.id

Article Info

History:
Submitted
February 11th, 2019

Revised
September 12th, 2020

Accepted
September 29th, 2020

Abstract

This study aims to produce an assessment instrument based on Higher-Order Thinking Skills for elementary school students. The assessment instrument was made in the form of a cognitive ability test for elementary school students of the sixth grade in Science subject at Elementary School. This research used research and development methods. The research design used is the Borg and Gall development model. The stages of development research include 1) Research and Information Gathering, 2) Planning, 3) Initial Product Development, 4) Initial Field Testing, 5) Main Product Revision, 6) Main Field Testing, 7) Operational Products Revision. The instruments used were interviews, product quality assessment sheets and questionnaires. The results of this research are a collection of valid, practical, and reliable HOTS-based assessment instruments. From the expert review results obtained an average assessment results percentage from the experts' validation of 93.57 with the category of very feasible. HOTS-based assessment instruments can be used in learning in the elementary schools.

Keywords: Assessment, HOTS, Elementary School

A. Introduction

Entering the industrial revolution 4.0 era and welcoming the era of society 5.0, human resources who are able to compete globally are needed. The education sector has a very important role in producing quality human resources. As contained in the Law of the Republic of Indonesia Number 20 of 2003 concerning the National Education System, Article 1 paragraph (1) states that education actively develops the students' potential to have religious-spiritual strength, self-control, personality, intelligence, noble character, and skills required of himself and society. Education means that education must be able to develop the students' potential in himself and its application in society, including in the world community.

The demands of 21st-Century life skills, the future generations to be able to live properly in the world community which consists of 16 skills which are categorized into three major categories, namely basic literacy skills, competency categories that are better known as 21st-Century Competencies, and character quality categories. The 21st-Century competency category consists of four skills (4C), namely critical thinking and problem-solving, creative thinking and innovation, communication, and collaboration. Right now is the industrial revolution 4.0 so the students must be able to adapt to the life challenges of the 21st-century. Moreover the Assessment and Teaching of 21st-Century Skills (ATC21S) categorizes the 21st-century skills into 4 categories, namely way of thinking, way of working, tools for working, and skills for living in the world (Griffin & Care, 2015). One of them in the way of thinking category includes creativity, innovation, critical thinking, problem-solving, and decision making. Other opinions expressed by Wagner (2009) that there are seven types of life skills needed in the 21st-Century , including the ability to think critically and solve problems. Based on the opinion of several experts, it can be concluded that critical thinking is one of the important elements of life skills that need to be mastered in the 21st-Century . (Wagner, 2009; Pellegrino & Hilton, 2013; Griffin & Care, 2015).

The importance of 21st-Century Skills so it needs to be developed from

an early age in students through education so on their productive age they can live properly in the world community.

Learning in schools is the spearhead of education implementation, one of them is in elementary schools. The learning process at school is largely determined by the quality of learning conducted by the teacher in the classroom. According to Hendracipta (2017), the critical thinking skills of elementary school students can also be improved by using the inquiry learning model. Apart from the learning model, the quality of learning can also be determined by the quality of the assessment. Assessment is a very important component in the learning process implementation.

Assessment of students in Elementary Schools refers to the assessment standards regulated in the 2016 Minister of Education and Culture Regulation No. 23, one of them aims to determine the achievement of students' competencies. Competencies to be achieved in learning as existed in the 2016 Minister of Education and Culture Regulation No. 20 explains that elementary school graduates must be able to think and act critically and creatively. Thinking and acting creatively are aspects of higher-order thinking skills.

Referring to the demands of the 2013 curriculum, it is necessary to develop higher-order thinking skills of students starting from elementary school. However, in reality, it has not been implemented in accordance with the demands of the 2013 curriculum. The results of the research of Rapih and Sutaryadi (2018) stated that 11.43% of teachers thought that higher-order thinking skills (HOTS) could not be started being taught to elementary school students. Moreover, according to Budiman and Jailani (2014), the teacher's ability to develop HOTS questions is still low and there is no specific assessment instrument for training HOTS. While according to Usmaedi (2017) HOTS should have started being taught and trained as early as possible. HOTS is a thinking ability that adjusts a person's cognitive level. Students at elementary school age must be introduced to HOTS according to their cognitive development at that age.

Higher-order thinking skills, when viewed from the cognitive domain in

Bloom's taxonomy revised, are at the level of analyzing, evaluating, and creating (Anderson & Krathwohl. 2001; Brookhart, 2010). Therefore, the assessment instrument used should not only measure low-level thinking skills but must be oriented towards higher-order thinking skills so it encourages students to develop their thinking skills. Students' thinking skills and ideas can be trained with the questions asked (Redhana, 2012). Other than that, according to Nada (2018), giving open questions can train students to think more broadly (creatively).

Reflection of the results of the Program Internationale for Student Assessment (PISA) in 2009 shows that the thinking ability of Indonesian students is still low. Indonesian students only master lessons up to level 3, while other countries have reached levels 4, 5, and even 6. In addition, the PISA results in 2000, Indonesia's ranking for Science was 38 out of 41 countries, in 2006 it was 52 out of 57 countries, 2009 was 61 out of 65 countries, 2012 was 64 out of 65 countries, 2015 was 62 out of 72 countries (OECD, 2016). The low PISA results are caused by many factors, one of them is that Indonesian students are generally less trained in solving problems with the characteristics such as the questions on PISA whose substance is contextual, according to reasoning, argumentation, and creativity in solving them. The results of the research by Aida et al, (2019) the critical thinking skills of elementary school students are still low and need to be trained using certain learning models. At a higher stage, college students' critical thinking skills are still deficient, especially in the evaluation and creation stages (Anugraheni, 2019).

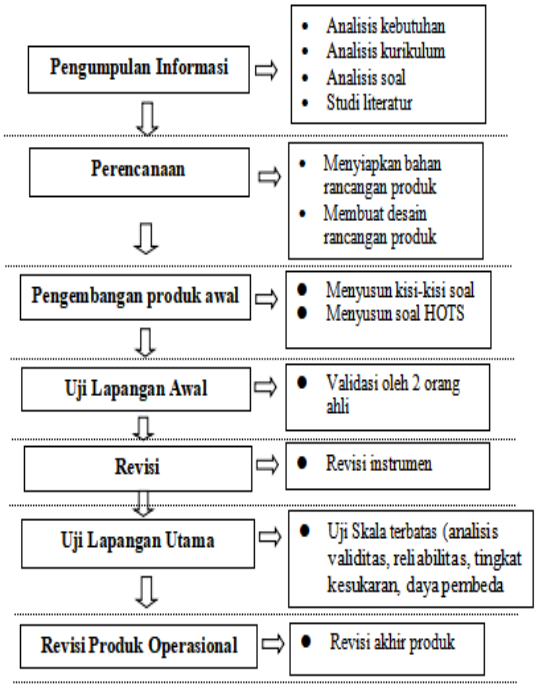
Given the importance of students' higher-order thinking skills, it is necessary to develop HOTS-based assessment instruments to measure students' higher-order thinking skills in elementary schools. Based on the background, the problem formulation in this research is "How to develop HOTS-based assessment instruments for elementary school students?"

The purpose of this research is to produce a HOTS-based assessment instrument that can be used in the learning process in elementary schools. The results of this research are expected to increase the insight and knowledge of

elementary school teachers in the preparation of HOTS-based questions, it can be used as an assessment instrument in learning in elementary schools, and are expected to improve the students' thinking quality by practicing doing the HOTS questions.

B. Research Methodology

The research method in this research using the method of Research And Development. The product developed is a HOTS-based assessment instrument for elementary schools. The stages used in this research refer to the Borg & Gall (1979) development model. This model consists of ten stages, namely: (1) information gathering, (2) planning, (3) initial product development, (4) initial field testing, (5) main product revision, (6) main field testing, (7) operational product revision, (8) field test, (9) final product revision, (10) dissemination and implementation. in this research will be carried out only until the seventh stage with modifications. Thus, the research procedure that was carried out according to the flow chart shown in Picture 1 below.



Picture 1. Flowchart of Research Activities HOTS-based Assessment Instrument Development for elementary students

a. Data Collection Technique

The data in the research were collected using these following instruments:

1. Interviews

Interviews were conducted for needs analysis which aims to obtain information about existing conditions

as a comparison or basic material for the instrument product being developed. Requirements analysis is the initial stage or preparation for development. Interviews were conducted with elementary school teachers who taught high-grade classes in the Palembang city.

2. Product assessment rubric

The assessment rubric that contains statements related to assessment instruments that can measure higher-order thinking skills to be assessed by experts who are experts in material and evaluation. This review is usually conducted before the questions are used or tested. The analysis of each item was conducted based on the writing rules, namely: examined in terms of material, construction, language/culture, the truth of the answer key/scoring guideline conducted by several reviewers. The experts' validation questionnaire is prepared based on a Likert scale with the alternative answers of very good (SB), good (B), sufficient (C), deficient (K), very poor (SK).

3. Product Test Results of Limited Scale

The product test results can be used to analyze the assessment instruments made. Whether an assessment instrument is good or not can be known after the analysis of the validity, reliability, difficulty level, and distinguishing power of the questions using ANATES V4 software. The test is conducted to a limited of 12 students.

b. Data Analysis Technique

The data obtained were analyzed quantitatively and qualitatively and then interpreted. The results of the research were discussed carefully. These results are compared with the results of other similar research and theoretical studies.

C. Research Result and Discussion

This development research uses the Borg and Gall model of development research. In this research, there are seven stages used to make HOTS questions so valid, practical, and reliable HOTS questions can be obtained which can then be used to measure the higher-order thinking skills of elementary school students. The seven stages are gathering information,

planning, initial product development, initial field testing, main product revisions, main field tests, and operational product revisions..

The first stage is to collect the initial data before designing a HOTS-based assessment instrument. The preliminary information collection was conducted by analyzing the teacher's understanding of HOTS-based assessment instruments in elementary schools. The analysis results of the teacher's understanding of HOTS-based instruments were carried out by distributing questionnaires to elementary school teachers in Palembang City. The results of the questionnaire data showed the understanding and application of HOTS-based assessment is still deficient. This is indicated by the results of the questionnaire as many as 50% of the teachers stated that they made their own HOTS-based assessment instruments and the other 50% of teachers stated that they used the HOTS cognitive level in the assessment. Teachers have difficulty in arranging the HOTS-based assessment instruments for learning in elementary schools. From the questionnaire results, 100% of the teachers stated that they had difficulty in arranging HOTS-based assessments. These results are in accordance with the findings of Mandra et al. (2019) that teachers do not understand the format and process of higher-order thinking skills-based assessments (HOTS).

Furthermore, entering the planning stage, at this stage a syllabus analysis is conducted in science subjects in sixth-grade elementary schools so the researchers can find out the material to be developed in the HOTS-based assessment instruments. Based on the results of this analysis, it is determined that the Basic Competencies that will be made into HOTS-based assessment instruments. Researchers limit the use of HOTS-based instruments for elementary school students in sixth-grade learning about the content of science learning in basic competencies of 3.1 Comparing how the plants and animals breeding; 3.2 relating the characteristics of puberty in boys and girls with reproductive health; 3.3 Analyzing how living things adapt to the environment.

To ensure the needs for HOTS-based instruments in elementary schools, a

material analysis was conducted. Material analysis is conducted to identify the main material that needs to be developed, collecting data, and select the material relevant to the development of the Higher Order Thinking Skills (HOTS) question instruments.

Table 1. Hots-Based Assessment Instruments List

No.	Kompetensi Dasar	Indikator	Materi	Tingkat Soal
1	3.1. Membanding-kan cara perkembang-biakan tumbuhan dan hewan	3.1.1. Mendeskripsikan tentang pelestarian makhluk hidup di lingkungan sekitar	Pelestarian makhluk hidup	C4
		3.1.2. Mengidentifikasi cara perkembangbiakan pada tumbuhan	Perkembangbiakan generatif dan vegetatif tumbuhan	C4
		3.1.3. Mengidentifikasi cara perkembangbiakan hewan	Perkembangbiakan generatif hewan	C4
		3.1.4. Mengelompokkan tumbuhan berdasarakan cara berkembangbiak	Tumbuhan di lingkungan sekitar	C4
		3.1.5. Mengelompokkan hewan di lingkungan sekitar berdasarkan cara berkembangbiak	Hewan di lingkungan sekitar	C4
2	3.2. Menghubungkan ciri pubertas pada laki-laki dan perempuan dengan kesehatan reproduksi	3.2.1. Menjelaskan masa pubertas	Pubertas	C4 C5
		3.2.2. Mengidentifikasi ciri perubahan pubertas pada laki-laki dan perempuan	Ciri-ciri pubertas pada laki-laki dan perempuan	C4 C4
		3.2.3. Menganalisis cara menyikapi pubertas	Sikap menghadapi masa pubertas	C5 C5
		3.2.4. Mengevaluasi upaya menjaga kesehatan reproduksi	Menjaga kesehatan reproduksi	C5 C5
		3.2.5 Menganalisis hubungan perubahan pada masa pubertas dengan kesehatan reproduksi	Hubungan pubertas dan kesehatan organ reproduksi	C5 C4
3	3.3. Menganalisis cara makhluk hidup menyesuaikan diri dengan lingkungan	3.3.1. Mengidentifikasi cara tumbuhan menyesuaikan diri dengan lingkungan	Cara beradaptasi tumbuhan	C4 C4
		3.3.2. Mengelompokkan tumbuhan berdasarkan cara menyesuaikan diri dengan lingkungan	Habitat tumbuhan	C4 C4
		3.3.3. Mengidentifikasi cara hewan menyesuaikan diri dengan lingkungannya	Cara beradaptasi hewan	C4
		3.3.4. Menganalisis hubungan adaptasi makhluk hidup dengan habitatnya	Adaptasi dan habitat	C5 C4
		3.3.5. Merencanakan cara melestarikan makhluk hidup	Pelestarian makhluk hidup	C6 C6

The results of the material analysis showed that several materials that are suitable for the development of the Higher Order Thinking Skills (HOTS) question instruments were presented in table 1. After drafting the draft, the next step was question production. At this question-making stage, researchers

designed the questions to measure higher-order thinking skills in science subject matter according to the list that had been made. This product design as a prototype. Each prototype focuses on three characteristics, namely material, construction, and language/culture. At this stage, the evaluation device is also prepared to be used to assess the HOTS question instrument that has been made in terms of material, construction, and language. According to Fanani (2013) HOTS assessment is the questions that generally measure the ability in the realm of analyzing, evaluating, and creating with the HOTS assessment characteristics, namely: measuring high-level thinking skills, based on contextual problems, not routine (not familiar), and using a variety of question forms.

The fourth stage is Preliminary Field Testing, at this stage, the test device that has been made are validated by experts. The results of the validation broadly include suggestions for improvements of less precise writing, conformity between items and learning indicators, the suitability of items with cognitive levels measured according to Bloom's taxonomy based on input and suggestions from the validators.

Table 2. Expert Validation Results on The Assessment Instruments

No	Aspek Penilaian	Hasil Penilaian		Rata-rata	Kategori
		Validator 1	Validator 2		
1	Materi	95,3	93,33	94.32	Sangat layak
2	Konstruksi	95,4	91,4	93,4	Sangat layak
3	Bahasa	96	90	93	Sangat layak
Total				93,57	Sangat layak

The sixth stage is the Main Field Testing. At this stage, the researchers conduct a limited scale trial of the instrument for assessing higher-order thinking skills. In this research, the product limited scale test was conducted at SDN 24 Palembang with 41 respondents.

The final stage is Operational Products Revision, Instruments that

have been tested then analyze the questions which including validity, reliability, difficulty level, distinguishing power of the item, using the software item analysis is conducted after the questions of HOTS prototype 2 tested, then calculated the validity and reliability scores. Each item will be analyzed the degree of difficulty and distinguishing power. This evaluation is conducted to find out whether the items that build the test have been able to perform its function properly or not.

1. The Level of Difficulty

To see the quality of the question items, not only validity and reliability requirements must be met, but calculations need to be done to determine the difficulty level of each item. A good question is a question that has a balanced level of difficulty (Arifin, 2009). The level of difficulty is needed to find out how difficult the questions being tested based on the test results done by students. Therefore, in the preparation of questions it is necessary to pay attention to the difficulty level.

From the results of field tests condicted on sixth-grade students of SDN 24 Palembang with 41 students as the respondents was obtained the calculation of the difficulty level. The results of the difficulty index calculation are interpreted in three criteria, namely: questions with P < 0,30 is a question that classified as too hard; questions with P 0.30 to 0.70 are classified as the questions with medium difficulty; and problems with P < 0,70 is classified as easy question. Based on the results of the item difficulty index analysis, there were 3 or 12% very hard questions, 6 items or 24% were hard questions, 16 questions or 64% were medium questions, and 0 or 0% were included in the easy questions.

According to Endrayanto and Harumurti (2014), a good question items are the items that are not too easy and not too difficult/hard. Problems that are too easy do not stimulate students to increase their efforts to solve them, otherwise, the questions that are too difficult/hard will cause students to become discouraged and have no enthusiasm to try again because they are out of their reach.

2. Distinguishing Power

The distinguishing power of a question is the ability of a question to

differentiate between high-ability students and low-ability students. The distinguishing power of the tests that have been developed has varied categories, namely very good, good, sufficient and poor. The quality of the distinguishing power can be determined by calculating the distinguishing index and then interpreting the results. According to Arifin (2009), The higher the coefficient of distinguishing power of an question item, the more able the question item is to differentiate between students who have mastered competencies and students who lack competencies.

The interpretation of the distiguishing power is based on the interpretation guidelines proposed by Arifin (2009). Question items that have a very good distinguishing power category have a distinguishing score in the range of >0.40. Question items that have a good distinguishing power category have a distinguishing power score in the range of 0.30 - 0.39. Question items that have a sufficient distinguishing power category have a distinguishing power score in the range of 0.20-0.29. Problems number 5 and 15 have a poor distinguishing power category that has a distinguishing power score in the range of <0.19. From the 25 question items made, there is 1 item that has a negative marked distinguishing power, namely number 16. A test question that has negative distinguishing power means that the question cannot differentiate the high-ability students from low-ability students. Question items that have negative distinguishing power should not be used in the next test because they have poor quality (Fatimah dan Alfat, 2019).

3. Reliability

A test can be said to be good as a measuring tool if it fulfill the requirements, one of them is the reliability. The reliability of the measuring instrument shows the extent of the measurement results with the tool can be trusted. This can be indicated by the level of consistency of the scores obtained by the subject as measured by the same measuring instrument under different conditions. (Solichin, 2017). The reliability score is obtained by using the internal consistency method, namely by dividing the test into two groups, the odd numbered question item group and the even numbered question

item group. Then, correlate the two groups. Based on the results of the ANATES V4 calculation, the overall reliability scre for the HOTS-based instrument developed was 0.73. The instrument results that were made as a benchmark for the reliability test were declared as reliable because they were greater than 0,70.

The reliability of this question is also influenced by the greater number of alternative answers to the multiple-choice objective test, the higher the test reliability (Alwi, 2010).

D. Conclusion

Based on the data that has been collected, it can be concluded that HOTS-Based Assessment Instruments in science subject content for sixth-grade are categorized as valid, reliable, and practical. The resulted instrument has a quality (feasibility) in the category of very feasible to use for learning in elementary schools. Suggestions for further research are to implement HOTS questions in the learning process that accordance with the characteristics of the 2013 curriculum.

References

Aida, T.N., , Subuh A, Ana Aa. 2019. Analisis Berpikir Kritis Siswa Melalui Model Poe (Predict-Observe-Explain) di Sekolah Dasar. *Jurnal Edukasia Elementaria*. *2*(2), 164–172.

Anugraheni, I. (2019). Analisis Kemampuan Berpikir Kritis Mahasiswa Dalam Menyelesaikan Permasalahan Bilangan Bulat Berbasis Media Realistik. *Scholaria: Jurnal Pendidikan Dan Kebudayaan*, *9*(3), 276–283.

Alwi, I. 2010. Pengaruh Jumlah Alternatif Jawaban Tes Obyektif Bentuk Pilihan Ganda Terhadap Reliabilitas, Tingkat Kesukaran Dan Daya Pembeda. *Jurnal Ilmiah Faktor Exacta*. 3(2). 184-193.

Anderson, L.W. & D.R. Kraathwohl. 2001. *A Taxonomy for Learning, Teaching, and Assesing*. New York: Longman Inc.

Arifin, Z. 2009. *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya.

Brookhart, S. M. 2010. *How to Assess Higher Order Thinking Skillss in Your Classroom.* Alexandria Virginia USA: ASCD.

Budiman & Jailani. 2014. Pengembangan instrumen Asesmen Higher Order Thinking Skill pada Mata Pelajaran Matematika SMP Kelas VIII semester 1. *Jurnal Riset*

Pendidikan Matematika. 1 (2). 139-151.

Borg, W. R. & M. D. Gall. 1979. *Educational research an introduction*. New York:longman.

Endrayanto & Harumurti. W. 2014. *Penilaian Belajar Siswa di Sekolah.* Yogyakarta: Kanisus.

Fanani Z.M. 2018. Strategi Pengembangan Soal *Higher Order Thinking Skill (HOTS)* dalam Kurikulum 2013. *Jurnal Edudeena*. 3 (1). 57-76.

Fatimah L.U dan Alfath K. 2019. Analisis Kesukaran Soal, Daya Pembeda Dan Fungsi Distraktor**.** *Jurnal Komunikasi dan Pendidikan Islam.*8 (2). 37-64.

Griffin, P., & Care, E. 2015. The ATC21S Method. In *Assessment and Teaching of 21st Century Skills.*

Hendracipta, N, Nulhakim, L, dan Agustin. 2017. Peningkatan Hasil Belajar Keterampilan Berpikir Kritis Pada Bidang Studi Ipa Melalui Metode Pembelajaran Berbasis Inkuiri. *Jurnal Pendidikan Sekolah Dasar*. 3(2). 215-227.

Kemdikbud. 2016. Permendikbud Nomor 20 Tahun 2016 tentang Standar Kompetensi Lulusan Pendidikan Dasar dan Menengah. Kemdikbud.

Mandra, S. Habib, S., N. 2019. Peningkatan Kualitas Pembelajaran Dan Penilaian Berbasis Hots.Jurnal hasil Pengamdian kepada masyarakat. 4 (2). 117-122.

Nada I, Utaminingsih S., Ardianti S,D. 2018. Penerapan Model *Open Ended Problems* Berbantuan CD Pembelajaran Untuk Meningkatkan Kemampuan Berpikir Kreatif Siswa Kelas Iv Sd 1 Golantepus. *Jurnal Pendidikan Sekolah Dasar*.4 (2). 216-227.

OECD. 2016. PISA 2015: *Result in Focus.* Paris: OECD Publishing.

Rapih & Sutaryadi. 2018. Perpektif guru sekolah dasar terhadap Higher Order Thinking Skills (HOTS): pemahaman, penerapan dan hambatan. *Premiere Educandum: Jurnal Pendidikan Dasar dan Pembelajaran.* 8 (1). 78 – 87.

Redhana I. W. 2012. Model Pembelajaran Berbasis Masalah dan Pertanyaan Socratik Untuk Meningkatkan Keterampilan Berpikir Kritis Siswa. *Jurnal Cakrawala Pendidikan*. 3. 351-365.

Pellegrino, J. W., & Hilton, M. L. (2013). Education for life and work: Developing transferable knowledge and skills in the 21st century. In *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*.

Solichin. M, 2017. Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes Dan Validitas Ramalan Dalam Evaluasi Pendidikan Jurnal manajemen Pendidikan Islam. 2 (2). 192-213.

Usmaedi, U. 2017. Menggagas Pembelajaran HOTS Pada Anak Usia Sekolah Dasar. *Jurnal Pendidikan Sekolah Dasar*. 3(1). 82-95.

Wagner, T. 2009. The Global Achievement Gap. *Physics World*, *22*(6), 52.