The International Journal of Social Sciences World
TIJOSSW is Available Online at:
https://www.growingscholar.org/journal/index.php/TIJOSSW
Vol. 3 No. 1, June 2021, pages: 165~175
DOI: https://doi.org
ISSN **2690-5167**
Growingscholar Publisher

Development of Assessment Instruments Based on Higher Order Thinking Skills in Thematic Learning

Ramadhania Giatri[1], **Rochmiyati**[2], **Pargito**[3]

Article history: Received March, 2, 2021;, Accepted: April, 5, 2021 , Displayed Online: May, 4, 2021; Published: June, 30, 2021

Keywords

HOTS;

Assessment;

Instruments;

Multiple choice test;

Thematic Learning;

Abstract

This study aims to develop an assessment instrument based on Higher Order Thinking Skills (HOTS) in thematic learning. This research adopts the Borg & Gall development model which is limited to eight stages of development, namely data collection, product draft planning, product development, validation, small group trials, small group trial results refinement, large group trials, and product improvement. This research was conducted on 110 grade 5 students in several elementary schools in Way Halim sub-district. This assessment instrument produced 20 multiple choice tests and was selected and tested twice. This shows that the assessment instrument is valid and reliable. The theoretical feasibility of the assessment instrument based on the assessment of 3 experts is included in the categories very good, relevant, representative, practical, discriminatory, specific, and proportional. Empirically this study shows that in the first and second trials the overall validity value of the questions is greater than the r table, namely 0.344 and the reliability values are 0.786 and 0.763. The results showed that the assessment instrument developed was theoretically and empirically feasible. So, efforts to support the learning process continue to be needed which are useful for improving higher order thinking skills in students.

[1] Postgraduate student, Lampung University, Lampung, Indonesia. Email: ramadhania.giatri.giatri92@gmail.com
[2] Senior Lecturer at Lampung University, Lampung, Indonesia. Email: rochmiyatiazwardi@yahoo.co.id
[3] Senior Lecturer at Lampung University, Lampung, Indonesia. Email: pargitos3@yahoo.co.id

1. Introduction

The implementation of the 2013 Curriculum is the government's effort to optimize and realize the goals of national education in Indonesia. Learning in elementary schools based on the 2013 Curriculum is integrated thematic learning for all grade levels. Learning and assessment are inseparable parts, because assessment is a major component in the duties and work of a teacher. If the teacher is able to do the assessment well, it is assumed that the teacher has good teaching ability as well. Assessment is a series of activities to obtain, analyze, and interpret data about the process and learning outcomes of students which are carried out systematically and continuously, so that it becomes objective information in decision making (Purwanti, et.al., 2008: 6)

The Indonesian government has formulated it in accordance with the Regulation of the Minister of Education and Culture No. 64 of 2013 concerning Basic and Secondary Education Content Standards to meet future needs and welcome the Golden Generation of Indonesia in 2045, has set Graduate Competency Standards based on XXI Century Competencies. Learning in the 21st century must be able to develop the competitive skills needed in the 21st century which focuses on developing Higher Order Thinking Skills.

According to Effendi, (2019: 41), basic thinking skills (Lower Order Thinking Skills) only use limited abilities on routine and mechanical things, for example memorizing and repeating information previously provided. Meanwhile, Higher Order Thinking Skills stimulate students to interpret, analyze or even be able to manipulate previous information so that it is not monotonous. Higher Order Thinking Skills (HOTS) or higher order thinking skills are skills to connect ideas and facts, analyze, explain, determine hypotheses to the conclusion stage. In accordance with the demands of the 2013 curriculum, students are not only able to know, understand and apply, but students are also required to be able to analyze, evaluate and even create.

Higher Order Thinking Skills is a thinking skill that not only requires memory skills, but also requires other higher skills. This is in line with the research results of Samritin and Suryanto (2016: 93) which explain that HOTS is very important because it can train students' abilities to process complex tasks or problems that involve connection, problem solving, and mathematical reasoning. Yee, et.al. (2015) state that HOTS learning makes students have analytical acuity, the ability to synthesize, and good evaluation skills of the problems they face. Indicators that measure Higher Order Thinking Skills include analytical skills (C4), evaluating (C5), and creating (C6).

A preliminary study of educators' opinions regarding HOTS instruments in Al-Azhar 1 Private SD, Al-Azhar 2 Private SD, Way Halim 1 Perumnas Public Elementary School, Way Halim Way Halim State Elementary School 2 and Way Halim Way Halim State Elementary School 3 involving 22 respondents, the results show that there were 64% of teachers who did not make their own assessment instruments but only relied on the instruments provided by the service, 55% of the teachers did not use a scientific approach in teaching which would support students' HOTS (Higher Order Thinking Skills) abilities.

Teachers play an important role in training students to have higher order thinking skills which are the demands of the 2013 curriculum. In order to have higher order thinking skills (HOTS), teachers can provide HOTS-based test questions to train students. Higher Order Thinking Skill (HOTS) based test questions can help students develop higher-order thinking skills.

The ability referred to is related to the ability to think critically, reflectively, metacognitively, and think creatively. This is in accordance with the objectives in the 2013 curriculum in PP. 17 of 2010, to prepare Indonesian people to have the ability to live as individuals and citizens who believe, be productive, creative, innovative, and affective and are able to contribute to the life of society, nation, state and world civilization.

The stimulus is the basis for the preparation of HOTS questions. The stimulus presented is contextual and interesting. The stimulus can come from global issues such as information technology, social, economic, health and education issues. Stimulus can also be raised from problems that exist in the environment around the educational unit such as culture, customs, cases in the region, or various advantages that exist in certain areas. The creativity of a teacher greatly affects the quality and variation of the stimulus used in writing HOTS questions.

Measuring the quality of a test can be done using qualitative analysis (theoretical) and quantitative analysis (empirical). Qualitatively, the test is said to be good if it meets the requirements of the construction, material, and language compilation. Meanwhile, the quantitative analysis was carried out by analyzing the items.

2. Materials and Methods

Assesment Instrument

It was also stated by Linn & Gronlund (Uno & Koni 2014: 1) that assessment is a general term that includes procedures used to obtain information about student learning (observation, average written test implementation) and learning progress assessment format. In line with this opinion, Kellaghan & Greany (2004: 5) state that "the term" assessment "will be used to refer to any producer or activity that is designed to collect information about the knowledge, attitudes, or skills of a leaner or group of leaners ". This means that the term "assessment" will be used to refer to any procedure or activity designed to gather information about the knowledge, attitudes, or skills of students or groups of students.

Higher Order Thinking Skills

Educating students by implementing the HOTS system means inviting students to think and get used to solving problems. Students are said to be able to think if they can apply their knowledge and develop their skills. HOTS is a higher way of thinking than memorizing facts, expressing opinions, or applying rules, formulas, and procedures in problem solving.

The characteristics of HOTS questions as expressed by Resnick (1987: 3) are as follows; (1) Non algorithmic. This means that the path of action is not completely predetermined. (2) Complex in nature. The path is not "visible" (mentally speaking) from every point of view. (3) Multiple solutions. Each with costs and benefits, not a unique solution. (4) Involves a variety of decision making. (5) Application of multiple criteria (many criteria) which sometimes contradicts one another. (6) It is effortful (requires a lot of effort). (7) There is a lot of mental work involved in the kind of elaboration and assessment required.

Multiple choice test

Planning for the development of higher order thinking skills based assessment instrument that will be used in this study, including formulating indicators of higher order thinking skills questions and formulating a grid for preparing higher order thinking skills questions. The form of tests used in educational units can be categorized into two, namely objective tests and non-objective tests. The objective tests that are often used are multiple choice forms, true false, matchmaking, and objective descriptions (Mardapi, 2012: 109). Reynoolds, et al. (2009: 182-184) the form of the test used can be divided into two, namely the selected response items and the constructed response items. The selected response items are divided into three, namely true-false items, multiple choice items, and matching items. Evaluation instruments that measure high-order thinking skills can use various types of assessment such as modified multiple choice, short answer construction, and long answer construction as has been done by Ramirez and Ganaden (2008). In this study the questions are used in the form of multiple choice objective questions (multiple choice test) which aims to train students' abilities in solving higher order thinking skills based questions.

Thematic Learning

Permendikbud Number 65 of 2013 concerning basic education process standards and states, that "In accordance with the Competency Standards for Graduates and Content Standards, the learning principles used are from partial learning to integrated learning". This was reaffirmed in Permendikbud Number 67 of 2013 concerning the Basic Framework and structure of the 2013 SD / MI Curriculum carried out through learning with an integrated thematic approach from grade I to grade VI. Joni (1996: 3) defines thematic learning as a learning system that allows students both individually and in groups to actively seek, explore and find scientific concepts and principles in a holistic, meaningful and authentic way. Thematic learning will occur when authentic events or theme exploration are the controllers in learning activities.

This is in accordance with Subroto's (2009: 9) opinion, thematic learning is learning that begins with a certain theme that links to other subjects, certain concepts are associated with other concepts that are carried out spontaneously or planned both in one or more fields of study and with various learning experiences so that learning becomes more meaningful.

According to Min, et al (2012; 274), explain the thematic approach as follows "According to the Malaysia Curriculum Development Center (2003), the strategic approach is an effort to integrate knowledge, skills and values learning and creative thinking using the theme. Teachers should encourage pupils to participate actively and physically in the processas a form of natural learning. Teachers should try to provide a meaningful learning experience to the students so that they do not only have fun but also show and display an interest in searching further from their own information. Students should also be given an opportunity to be independent, explore and experience learning themselves. Thematic learning process will help students to think creatively and critically. Thematic approach is the meaningful learning for students because they learn to do independently ".

Time and type research

This research is began in July 2020 using quantitative approach. The type of research used in this study is Research and Development (R & D) methods. According to Sugiono (2013, p. 407) R&D

research method is a development research method used in producing a product. This development research was adapted based on the development model of Borg & Gall which consisted of 10 steps of development. But in this study researchers only used 8 steps of development, namely data collection, product draft planning, product development, validation, small group trials, small group trial results refinement, large group trials, and product improvement. in the form of questions based on higher order thinking skills thematic learning.

Research Subjects

Subjects of this research are 110 grade 5 students in several elementary schools in Way Halim sub-district. The students filled out a test which was given online via google form. The technique of determining the number os samples in this research uses cluster sampling. According to Sugiyono (2012: 121) regional sampling techniques are used to determine the sample if the object to be studied or the data source is very broad. Thus, the researcher gives equal rights to each subject to have the opportunity to be selected as a sample.

Procedures

The non-test technique is used to obtain qualitative data by reviewing the assessment instrument by experts in the form of a questionnaire. The questionnaire in this study consisted of three, namely: a needs analysis questionnaire during field observations, an expert validation questionnaire, a practitioner / educator response questionnaire. The data that will be processed is data in the form of comments, suggestions, and product improvements from expert validation. The questionnaire for the assessment instrument sheet is assessed by giving a check mark (√) according to the indicators in the item questions.

The research data collection was carried out through several stages, there are; (1) compiling a list of questions on the Google form for students (2) distributing questions via email and WhatsApp to students (3) instruments distributed in the form of multiple choice assessment instruments (4) summarizing and analyzing data according to the responses of learning outcomes that have been carried out by students during research.

Instrument

The instrument of this research is the assesment used is a multiple choice objective test with four alternative answers. One item of multiple choice type consists of two parts, namely: a statement or stem, and an alternative answer or option. In addition to multiple choice, this study developed a multiple choice test. However, previously, the product prototypes were then assessed and evaluated by experts using a questionnaire to see the instrument's suitability. The validator consists of evaluation experts, material development experts, and linguists. Suggestions from experts were used to revise Prototype I. Feasibility analysis was obtained using the following formula.

$$Final\ score = \frac{Score\ obtained}{Maximum\ Score} \times 100$$

Final scores were converted into the following categories as shown in Table 1 below.

Table 1.
Criteria for evaluation of items by Experts

Interval	Score Description
76 – 100	Very Good
51 – 75	Good
26 – 50	Good Enough
0 – 25	Not Good

(Arifin and Retnawati, 2017:101)

3. Results and Discussions

Based on the research results, it was found that the items made by the educators did not match the demands of the 21st century. Educators do not review the material before the exam. Educators do not make higher-order thinking assessment instruments and do not analyze items. However, theoretically, the assessment needs to be done by educators to measure the extent to which students understand the learning material delivered by educators (Hosnan, 2014), where the results of the assessment can be used to determine students' competencies or abilities and their learning achievement (Kankam Boadu, et al., 2015)).

The findings of this study are in line with the findings of Nova, et. al., 2016; Budiman & Jaelani, 2014 that the assessment instrument needs to be tested to obtain theoretical and empirical feasibility. The theoretical feasibility test was carried out by three experts, namely evaluation experts, material development experts, and linguists. To determine the empirical feasibility it was tested on students where the test results were then analyzed to determine the validity, reliability, difficulty level, and distinguishing power. Novitasari N. et.al, (2015) also explained that the assessment instrument needs to be tested to obtain theoretical and empirical feasibility results. The development of test instruments in this study refers to Borg & Gall (1983) as following steps.

3.1 Needs Analysis and Problem Identification

Several problems such as conventional questions made by educators, lack of understanding of educators in making higher order thinking assessment instruments, analysis of teacher problem item deficiencies were identified. However, there are some potential supports for conducting this research such as support from educators and representatives of the identification of school facilities and infrastructure.

3.2 Data Collection and Product Planning

Deeper analysis of the 2013 curriculum applied in schools is carried out carefully followed by analyzing core competencies and basic competencies.

Table 2. Core Competencies and Basic competencies

Core Compentencies	3. Understanding factual knowledge by observing (listening, seeing, reading) and asking questions based on curiosity about

	himself, God's creatures and their activities, as well as objects they encounter at home and at school
Basic Competencies	3.3 Examining the socio-cultural diversity of the community (Education and Citizenship). 3.5 Extracting important information from historical narrative texts presented orally and in writing using the following aspects: what, where, when, who, why, and how (Indonesian Subject). 3.7 Analyzing the effect of heat on changes in temperature and shape of objects in everyday life (Natural Science). 3.4 Identifying the important factors that caused the colonization of the Indonesian nation and the efforts of the Indonesian nation to defend its sovereignty (Social Science). 3.2 Understanding scales (Art and Culture).

After that, it was continued with the making of higher-order thinking instruments, selecting stimuli, making questions, making answer keys, and preparing assessment guidelines.

3.3 Initial Product Design

The initial product design or prototype is made at this stage. Done based on the concept of critical thinking by referring to the assessment instrument made. Later, Prototype I was developed.

3.4 Product Design Validation

At this stage the product prototype is assessed and evaluated by experts using a questionnaire to see the suitability of the instrument. Validators consist of evaluation experts, material development experts, and linguists.
Suggestions from experts were then used to revise Prototype I and to state that the design of test instruments was
feasible.

Table 3. Expert's Validation

Numb.	Suggestion of Expert's Evaluation	Revision Results
1.	The subject matter is formulated briefly, clearly and firmly.	As Suggested
2.	The formulation of the subject matter and the choice of answers are just statements that are needed.	As Suggested
3.	The subject matter does not provide an answer key clue.	As Suggested
4.	Free subject matter and statements that are double negative.	As Suggested
5.	The choice of answer is homogeneous and logical in terms of material.	As Suggested
6.	Pictures, graphs, tables, diagrams or the like are clear and	As Suggested

	functional.	
7.	The length of the answer choices is relatively the same.	As Suggested
8.	The answer choices do not use the statement "all the above answers are wrong / true" and the like.	As Suggested
9.	The answer choices in the form of numbers / times are arranged according to the order of the size of the numbers or chronology.	As Suggested
10.	The item of the question does not depend on the answer to the previous question.	As Suggested
11.	Problem according to the characteristics of Hihger Order Thinking Skills (HOTS).	As Suggested
Numb.	**Suggestion of Expert's Material**	Revision Results
1.	Questions according to Basic Competencies	As Suggested
2.	Questions according to the indicators (requires a multiple choice written test).	As Suggested
3.	The material asked is in accordance with the competence (urgency, relevance, continuity, high daily usability).	As Suggested
4.	Problem according to the characteristics of Higher Order Thinking Skills (HOTS).	As Suggested
5.	Homogeneous and logical answer choices.	As Suggested
6.	There is only one answer key.	As Suggested
Numb.	**Suggestion of Linguist**	Revision Results
1.	Formulation of communicative question sentences.	As Suggested
2.	Do not use words / expressions that cause multiple interpretations / misunderstandings.	As Suggested
3.	Do not use the local language / taboo.	As Suggested
4.	The answer choices do not repeat the same word / group of words, unless they constitute a unified meaning.	As Suggested
5.	Guided by the rules of writing standard questions from various forms of assessment questions.	As Suggested

Expert's validation results were included in the excellent categor y as shown in Table 4 below.

Table 4. Expert Validation Results

Numb.	Expert	Score 1	Score 2	Average	Category
1.	Evaluation	68,18	86,36	77,27	Good
2.	Material	70,83	79,13	75	Very Good
3.	Language	75	85	80	Very Good

3.5 Product Revision

Product revisions were based on expert advice. After the changes were made based on expert feedback, Prototype II was then developed followed by a main field test in a small class. The test results at this stage were as follows.

3.6 Small Group Test

Description Test

Table 5. Value of Instruments

Number of Question	Amount	Description
1, 2, 4, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 21, 24, 25, 27, 28, and 29	20	Valid r count > r table
3, 5, 6, 7, 16, 20, 22, 23, 26, and 30	10	Invalid r count < r table

Table 6. Instrument Reliability

Number of Question	**R value count**	**Criteria**
1 – 30	0,954	Very High

Table 7. Question's Difficulty Levels

Category	**Question**	**Amount**
0,71 – 1,00 (Easy)	2, 5, 11, 12, 14, 15, 19, 22 and 25	8
0,31 – 0,71 (Moderate)	1, 3, 4, 6, 7, 8, 9, 10, 13, 16, 17, 18, 20, 21, 22, 23, 24, 26, 27,28, 29, and 30	22
0,00 – 0,31 (Difficult)	-	-

Table 8. The Differences of Question Levels

Distinguishing Index	**Number of Question**	**Amount**
0,40 – 1,00	6 and 14	Very Good
0,30 – 0,39	1, 4, 7, 8, 10, 12, 14, 15, 16, 17, 18, 21, 24, 26, 27, 28, and 29	Good
0,20 – 0,29	2, 3, 9, 11, 19,20, 22, 25, and 30	Good Enough
0,00 – 0,19	-	Not Good

Each item in Prototype II was analyzed. The ten description questions were considered empirically feasible because they prove to be valid and reliable with moderate levels of difficulty and good differentiation. The effectiveness of liking was also proven to be good. After Prototype II was revised, Prototype III was developed which was then tested in a larger size class.

3.7 Large Class Instrument Trial Results

Prototype III was tested for four subjects. The test results were analyzed to determine the validity, reliability, level of difficulty, and distinguishing features of each item. The following were the results of the analysis.

Table 9. Value of Instruments

Question	**Amount**	**Description**

1-20	20	Valid r count > r table
-	-	Invalid r count < r table

Table 10. Instrument Reliability

Number of Question	R value count	Criteria
1 – 20	0,839	Sangat Tinggi

Table 11. Question's Difficulty Levels

Category	Question	Amount
0,71 – 1,00 (Easy)	1, 6, 19, and 20	4
0,31 – 0,71 (Moderate)	2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, and 18	16
0,00 – 0,31 (Difficult)	-	

Table 12. The Differences of Question Levels

Distinguishing Index	Number of Question	Amount
0,40 – 1,00	4 and 14	Very Good
0,30 – 0,39	1, 3, 5, 7, 8, 9, 10, 11, 13, 15, 16, 17, 18, and 19	Good
0,20 – 0,29	2, 6, 12, and 20	Good Enough
0,00 – 0,19	-	Not Good

Each item in Prototype III was analyzed. The ten descriptive questions were considered empirically feasible because they prove to be valid and reliable with moderate difficulty and good differentiation. After Prototype III was revised, Prototype IV as the final product was developed.

4. Conclusion

Based on the results and discussion, it can be concluded that the final product in this study is a HOTS-based assessment instrument which is theoretically and empirically feasible for thematic learning of fifth grade elementary school students. The suitability of the instruments was obtained from the evaluation and testing of experts in the classroom. This instrument is theoretically feasible because it has been validated by experts in the fields of evaluation, material, and language, where the results are included in the very good category. This assessment instrument is empirically feasible because it is tested in the classroom. The test results proved to be valid and very reliable with moderate difficulty and good discrimination.

Acknowledgements

Thank you to SD Al-Azhar 1 Bandar Lampung, SD Al-Azhar 2 Bandar Lampung, SDN 1 Perumnas Way Halim, SDN 2 Perumnas Way Halim, and SDN 3 Perumnas Way Halim for giving me permission to do research. Thank you for all teacher grade 5 for always helping, support, and listen to me during my research.

References

Effendi, Gunarto Wahid, (2019). Pelatihan Pembelajaran dan Penilaian Berbasis HOTS (Higher Order Thinking Skill) bagi Guru SD. *Jurnal Indonesia Mengabdi. Vol 1, No. 2, 40-42.*

Joni, R.T. (1996). *Pembelajaran Terpadu*. Depdikbud. Dirjen, Dikti, Proyek Pengembangan Guru Sekolah Dasar. IBRD LOAN 3496-IND: Jakarta.

Kankam, B. et al. (2015). Persepsi guru tentang praktik teknik penilaian otentik dalam pelajaran IPS di sekolah menengah atas di Ghana. Jurnal Internasional Penelitian Pendidikan dan Ilmu Informasi. 10 Januari 2015. *1* (4), 62-68.

Kellaghan,T.,& Greaney, V. (2004). *Assesing Student Learning in Africa*. Clearance Centre, Inc. Washington.

Mardapi, (2012). *Pengukuran Penilaian & Evaluasi Pendidikan*. Nuha Medika: Yogyakarta.

Min, K.C., Rashid,M.A, & Nazri,M.I., (2012). Teachers' Understanding and Practice towards Tematics Approach in Teaching Integrated Living Skills (ILS) in Malaysia. *International Journal of Humanities and Social Science*, Vol. 2 No. 23:273-281. Diakses pada 10 Juli 2017.

Nova, AR, et.al. (2016). Pengembangan instrumen penilaian konsep teslet pada materi suhu dan kalor. Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan. *1* (6), 1197-1203.

Novitasari, et.al. (2015). Mengukur keterampilan memecahkan masalah siswa sekolah menengah tentang biologi. Jurnal Biologi Edukasi. *7* (1), 1-6.

Purwanti, et, al. (2008). *Assesmen Pembelajaran SD*. Depdiknas: Jakarta.

Retnawati, H. (2014). *Teori Respons Butir dan Penerapannya untuk Peneliti, Praktisi Pengukuran dan Pengujian Mahasiswa Pascasarjana*. Yogyakarta: Nuha Medika.

Resnick, L. B. (1987). The 1987 presidential address learning in school and out. *Educational researcher*, *16*(9), 13-54.

Samritin & Suryanto. (2016). Developing An Assessment Instrument Of Junior High School Students' Higher Order Thinking Skills In Mathematics*. Research and Evaluation in Education*. Vol. 2(1):92-107.

Subroto, S. (2009). *Proses Belajar Mengajar di Sekolah*. Rineka Cipta: Jakarta.

Sugiyono. (2010). Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualititatif, dan R&D. Bandung: Alfabeta.

Uno, Hamzah B. dan Koni, Satria. (2014). *Assessment Pembelajaran: Salah satu bagian penting dalam pembelajaran yang tidak dapat diabaikan adalah pelaksanaan penilaian.* PT Bumi Aksara: Jakarta.

Yee, M. H., Yunos, J. M., Othman, W., Hassan, R., Tee, T. K., & Mohamad, M. M. (2015). Disparity of learning styles and higher order thinking skills among technical students. *Procedia-Social and Behavioral Sciences*, *204*, 143-152.