TÜRK FEN EĞİTİMİ DERGİSİ
Yıl 17, Sayı 4, Aralık 2020

Journal of
TURKISH SCIENCE EDUCATION
Volume 17, Issue 4, December 2020

http://www.tused.org

The Analysis of Instrument Quality to Measure the Students' Higher Order Thinking Skill in Physics Learning

Edy SUPRAPTO[1], SARYANTO[2], Rudy SUMIHARSONO[3], Syahrul RAMADHAN[4]

[1]Mechanical Engineering Education Study Program, Faculty of Teacher Training and Educational Sciences, The University of Nusa Cendana, Kupang, Indonesia, ORCID ID: 0000-0003-4796-5080
[2]Mathematics Study Program, Faculty of Teacher Training and Education, The University of Alma Ata Yogyakarta, Indonesia, ORCID ID: 0000-0003-4632-2358
[3]IKIP PGRI Jember, JL. Jawa 10 Jember, Indonesia, ORCID ID: 0000-0001-8910-6079
[4]Universitas Negeri Yogyakarta. Jalan Colombo No. 1, Karangmalang, Yogyakarta 55281, Indonesia, ORCID ID: 0000-0002-2661-2537

Received: 27.03.2019 **Revised:** 30.04.2020 **Accepted:** 08.10.2020

The original language of article is English (v.17, n.4, December 2020, pp.520-527, doi: 10.36681/tused.2020.42)

Reference: Suprapto, E., Saryanto, Sumiharsono, R. & Ramadhan, S. (2020). The Analysis of Instrument Quality to Measure the Students' Higher Order Thinking Skill in Physics Learning. *Journal of Turkish Science Education,* 17 (4), 520-527.

ABSTRACT

This research aims to produce feasible and valid assessment instrument of Higher Order Thinking Skill (HOTS) to measure students' Higher Order Thinking Skill in Physics learning. The type of this research was research and development, adapted from development model from Brog and Gall. The researchers modified Borg and Gall's development model as stages such as (1) Collecting information, (2) Making the plan, (3) Preparing the product form, (4) Conducting revision of the initial product, (5) Implementing the product. The study group consisted of 34 10th grade students of Senior High School 1 Sape. Data collection was conducted through a test consisted of 20 multiple-choice items of Physics learning in the form of HOTS. For data analysis, the technique of QUEST was used to examine the validity, reliability, the difficulty level, and distinguished items. This research generated an assessment instrument that is feasible in aspects of validity, reliability, difficulty level, and question differentiaton to be used as an alternative instrument in assessing students' HOTS.

Keywords: Higher order thinking skill, physics, instrument.

INTRODUCTION

Nowadays, there are many students who are considered to have no higher order thinking skills (HOTS), so it becomes the background for the writer to develop a HOTS assessment instrument in order to reform the inappropriate learning systems. At first, HOTS emerged from Bloom's (1956) concept in a book titled "Taxonomy of Educational Objectives: the

Correspondence author e-mail: syahrul.ramadhan2015@student.uny.ac.id.

classification of Educational Goals" that agglomerates various thinking levels of Bloom's Taxonomy, from the lowest to the highest thinking level.

Bloom's concept of the learning goal is divided into three aspects as Cognitive (mental action to acquire knowledge), Affective (emotional side about attitudes and feelings, and Psychomotor (physical ability, such as ability to perform) (Anderson et al., 2001). Regarding the efforts of education advancement, higher order thinking skills (HOTS) in learning and teaching process are indispensable in accordance with the current development and evolutionary demand of education (Ramadhan, Mardapi, Prasetyo, & Utomo, 2019). HOTS is a tool to facilitate thinking process with many variables in particular conditions. Basically, the most found case is on the importance of guidance and encouragement for students to achieve HOTS goals (Ramadhan, Sumiharsono, Mardapi, & Prasetyo, 2020).

The teacher or educator has a role to be fully responsible in learning Physics at schools. Educators as the main source of information, using recitation as teaching methods based on books and lectures (Batlolona, Diantoro, Wartono, & Latifah, 2019). At the end of the day, the students seem to lack enthusiasm, and as the long time passes, get sleepy and lose focus. Students should get involved in many activities that are supposed to be done by teacher to have conducive learning atmosphere, so that they are involved in activity and creativity (Nguyễn & Nguyễn, 2017). The main problem is the way to deal with learning process that is not only focused on the teacher as a transformer, but also involves many development processes to make students better at analyzing, evaluating, and creating in each lesson.

The descriptions above are the background of the writer to develop Higher Order Thinking Skills assessment instrument, especially the feasibility and validity of multiple-choice HOTS items. The basic competency used is KD 3.6 "The Implementation of Newton's Laws and Concepts" in physics 10th grade for senior high school students. Moreover, the writer expects that this study can help students in analyzing, evaluating, and creating sophisticated and progressive thinking transformation that is beneficial to society in the future.

LITERATURE REVIEW

Assessment Instrument

The instrument is a tool that meets academic requirements that measures an object to know about what is accurately obtained through a valid and reliable method; and assessment means a process of obtaining information to make decisions about students, curriculum, programs and educational policies (Satria & Uno, 2012). Mardapi (2008) said that assessment instruments have two types, test and non-test. In the framework of measurement and assessment, tests are for measuring learning achievement, intelligence, talents, and skills of students, while non-tests measure assessments, attitudes, observation guidelines, etc. (Mardapi, 2004, 2008). The assessment has a function as (1) a tool used to find out whether or not the instructional objectives have been achieved. The assessment must refer to instructional formulations through this function; (2) a tool as feedback to improve the teaching and learning process. The improvements might be implemented in instructional activities, student learning activities, teacher teaching strategies and others; and (3) report of learning progress in many science subjects to student's parents.

Higher Order Thinking Skill (HOTS)

Higher Order Thinking Skills explained by King, Goodson, and Rohani (1998) is a process of selective, creative, logical, critical, and meta-cognitive thinking. The implementation of thinking concept when students have difficulties. The aspects of HOTS are identified by Brookhart (2010) as analysis, evaluation and creation, reason in logic, ability to think critically, ability of problems solving and thinking creatively.

The thinking concept developed by Bloom is called Bloom's Taxonomy. Bloom explains that there are two cognitive processes namely HOTS skills that involve synthesis and analysis (C4), evaluating (C5) and creating or creativity (C6) also low order thinking skills or LOTS which involve the ability to recitation (C1), understanding (C2), and implementing (C3) (Anderson et al., 2001: Ong, Hart, & Chen, 2016).

In line with Bloom's taxonomy, Zohar & Cohen (2016) explains that the ability of recitation (C1) is limited to repeating past events; understanding (C2), includes absorption of information, interpreting meaning, and exploring; implementing (C3) is to generalize a situation that has been described previously; analyzing (C4), connects knowledge with one another systematically and in a structured manner and ability of problem-solving through facts; evaluating (C5) means conducting an assessment based on criteria or standards; and creating (C6) as the highest level of HOTS is where students can have the ability of problem solving through creative thinking.

Multiple-choice

Multiple-choice tests are objective in the large-scale and small-scale of test (for example, the Formative Test, the Summative Test). Multiple choice questions are fairly easy in scoring. Gronlund and Linn (1990) revealed that to measure simple thinking to complex thinking, multiple choice questions adjusted to the subject matter can be used. Multiple-choice has a parameter that causes a high level of difficulty, such as the existence of distractor to deceive the answer.

HOTS issues generally prioritize the insertion of stimulus in contextual situations (Abdurrahman, Setyaningsih, & Jalmo, 2019; Tiruneh, De Cock, & Elen, 2017). The answer key is not explicitly contained in the reading or stimulus. Respondents can find answers to questions about reading using the knowledge background, and state the reasons. The complexity of multiple-choice questions is to test students' understanding of a problem comprehensively related to one statement to another. Similar with multiple-choice questions, HOTS questions in the form of multiple-choices include stimulus based on the contextual situations (Jensen, McDaniel, Woodard, & Kummer, 2014).

In Indonesia, Higher Order Thinking skill is always linked with Bloom’s taxonomy revised, mainly top three levels as C4 (analyzing), C5 (evaluating), C6 (creating). Even the use of Bloom’s taxonomy is also loaded in the curriculum used in Indonesia. Therefore, the concept of Higher Order Thinking skill used in this study refers to the idea of higher-order thinking from Bloom’s taxonomy revised.

METHODS

This study as a Research and Development study, has an objective to develop a HOTS scale for physics students of 10th grade students of Senior High School 1 Sape. The procedure in this development was adapted from development steps by Gall, et al. (1996). The writer modified the steps of research development by Borg and Gall (1996) into several stages, (1) Gathering information; (2) Making a plan; (3) Preparing product forms; (4) Conducting initial product revisions; (5) Product implementation.

The HOTS assessment instrument in this study used 20 physics questions based on KD (Basic Competencies) in multiple choice. The test instrument was distributed to 34 respondents using learning materials "The implementation of Newton’s laws and concepts” that had been studied previously and also about contextual cases. The descriptive analysis techniques are used to process the data that was conducted from the results of limited trials in the field. The formula used to measure this percentage is the formula using the QUEST program. The QUEST program is used to measure the validity, reliability, level of difficulty

and differentiation of items. The validity results were conducted through MNSQ INFIT analysis and Item fit. The instrument reliability test in this study using Rasch model which facilitates the interpretation of statistical reliability test results. The distinguishment power analysis uses the biserial point value in the Quest program.

RESULTS

There are two stages that must be done before starting the test. At the first stage, the instrument was assessed by several experts, consisting of 1 instrument expert, 1 product expert, and 2 material experts. The second stage, 34 students of 10th grade students of Senior High School 1 Sape participated in a trial test on multiple choice HOTS questions that had passed the expert validation test.

The Quest Program is an item analysis application developed based on applied statistics based on a theory namely item response. Modern measurement theory is used in item analysis. Latent Trait Theory (LTT) or Characteristics Curve Theory (CCT) is another name for item response theory. There are two postulates as the basic of item response theory. The first one is a set of factors namely traits, latent traits or abilities that can predict the ability of the subject. Verbal abilities, psychomotor abilities, cognitive abilities are called traits. The second postulate is the item characteristics curve (ICC) which has the latent ability of respondents and item sets. The logistics model is studied in PMM activities, named a one-parameter logistic model (rasch model) or 1-parameter logistic response theory (IRT 1-PL) to analyse the data that focuses on the level of difficulty parameters.

Adams and Khoo (1996) stated that Quest can analyse items. The Rasch Model is a central element, one parameter (1-PL). The Quest Program is a participant's ability = Θ□and the difficulty level of item b as the main item. Itanal in the syntax section is output command on the statistics of test on difficulty level, discrimination level, and distractor level. The output provides information about item statistics and test kits such as the degree of difficulty and discriminatory power. Quest analyses respondents who are judged dichotomously (1-10) or politically (1-2-3-4-etc.). Unconditional (UCON) or joint maximum like hood is used by Quest to estimate the subject. The Quest program is used to be able to measure the validity, reliability, level of difficulty and differentiation of questions.

Results of Limited Test Data Validity

The good learning outcomes are valid results tests (Ramadhan et al., 2019; Ramadhan et al., 2020). Limited trials were conducted in 10th grade students of Senior High School 1 Sape involving 34 students (one class). The multiple choice HOTS question in a limited trial is conducted in 60 minutes and one trial only.

The validity results were obtained through MNSQ INFIT analysis and Item fit. The problem is declared valid if it is in the range of -2.0 to +2.0 with the FIT statement. After analysis results, 20 items were declared fit. Here are the results of the validity of the questions using INFIT analysis of MNSQ data from 34 students of 10th grade students of Senior High School 1 Sape.

Table 1. *Problem multiple choice HOTS declared Valid*

No Item	INFIT MNSQ	Criteria
1	1,03	FIT
2	0,99	FIT
3	0,84	FIT
4	1,07	FIT
5	0,88	FIT

No Item	INFIT MNSQ	Criteria
6	1,12	FIT
7	0,99	FIT
8	1,15	FIT
9	0,92	FIT
10	1,00	FIT
11	0,96	FIT
12	0,96	FIT
13	0,98	FIT
14	1,02	FIT
15	1,05	FIT
16	0,87	FIT
17	0,85	FIT
18	0,99	FIT
19	1,23	FIT
20	0,98	FIT

Item Reliability Analysis Problem

Reliability is a measuring tool to determine a quality of item. A test is reliable if it is tested on the same group at different times. The measurement of reliability tested in many conditions and opportunities must have the same result (Mardapi, 2008).

The analysis of Items fit if it is in the range of 0.77 to 1.30 then items are considered to be valid. The questions made by writer were valid. The reliability value shows that the questions reliability in high category that is 0.87. It means that the test instrument is reliable, but it is still not very good due to its high level of the reliability coefficient of education. The average level of compatibility of the items is 1.0 and the standard deviation is 1.11, so overall the respondents are suitable with the model set of Rasch.

Item Difficulty Level Analysis

Boopathiraj and Chellamani (2013) define that item difficulty is the proportion of respondents who marks the items correctly. Items with medium difficulty and not easily answered are good questions.

Table 2. *Results of difficulty level output of the Quest program*

Item	***Threshold***	Criteria
1	-1,40	Difficult
2	1,41	Easy
3	1,41	Easy
4	0,18	Difficult
5	1,10	Easy
6	-0,73	Difficult
7	-0,19	Difficult
8	0,81	Easy
9	-0,88	Difficult
10	-1,21	Difficult
11	1,41	Easy
12	1,59	Easy
13	0,55	Medium

Item	*Threshold*	Criteria
14	-0,88	Difficult
15	-1,21	Difficult
16	0,81	Easy
17	0,81	Easy
18	-1,04	Difficult
19	0,06	Difficult
20	-2,61	Difficult

Based on the data in table 2 above, 11 problems in difficult level are 55% of the test. Questions with an easy level are 40% were 8 items, and 5% of the questions are medium level of difficulty was found only 1.

The Analysis of Distinguished Items

According to Mardapi (2008), whether an item is able to distinguish students who have low or high ability is one of problem analysis objectives. The feature of its ability to categorize is that it has a positive discrimination index. Students in this category are smart students. Students in the smart category answer more questions correctly. The item is said to have no distinguishing ability at all with symbol D = 0. It means that both of the Upper group students and Lower group students answered correctly.

Table 3. *Results of distinguishing power using biserial points*

Item	*Point Biserial* (ρ_{bis})	Criteria
1	0,18	Bad
2	0,24	Enough
3	0,53	Good
4	0,20	Enough
5	0,49	Good
6	0,08	Bad
7	0,26	Bad
8	0,08	Bad
9	0,40	Good
10	0,28	Enough
11	0,30	Good
12	0,33	Good
13	0,35	Good
14	0,18	Bad
15	0,10	Bad
16	0,49	Good
17	0,52	Good
18	0,28	Enough
19	-0,06	Bad
20	0,22	Enough

The data in table 2 shows 8 good quality questions, 7 poor quality questions and 5 mediocre quality questions. It means that the questions made by writer is accepted because the majority of questions are acceptable and can be implemented on students.

Product Revision

The validity and reliability criteria are conducted to gain the final product in product revision. The validation revision and product revision on limited trial are the product revision of this study that based on product trial assessment. The average HOTS test questions on the Basic Competence "The Implementation of Newton's Laws and Concepts" which consists of 20 feasible and valid HOTS questions. Generally, the insights and suggestions from the validation process helped a better version are the language, produce questions, material focus, and material sequence.

Final Product Review

The HOTS assessment tool for the basic competence "The Implementation of Newton's laws and concepts" in physics for 10th grade students of Senior High School 1 Sape is the final result of this study. HOTS multiple-choice questions developed have been observed as a valid and reliable instrument in limited trials. Instrument experts, product experts and material experts are involved in the process of perfecting this product. The improvements were made after getting the results from validator validation and limited trials. The product developed has met the criteria of a decent item. The quality of items has been tested through validation, reliability, level of difficulty and distinguishing features.

CONCLUSION

Based on the results of research and discussion, the conclusions are (1) the findings about assessment instrument of Higher Order Thinking Skills (HOTS) on Basic Competence "The Implementation of Newton's Laws and Concepts" 10th grade students of Senior High School 1 Sape. The multiple-choice of HOTS as instrument provided with five options; (2) The validity of HOTS questions based on validator analysis including the instrument experts, the product experts, and the material experts. The result of instrument expert analysis shows that HOTS assessment instrument is feasible from the aspect of validity, reliability, difficulty level, and question differentiation to be used as alternative questions that will be tested at schools and (3) Characteristics of multiple choice HOTS questions show that the quality of the question item was obtained from the result of question item analysis. The calculation result of HOTS questions validity showed that all 20 questions were valid.

ACKNOWLEDGEMENTS

This work was supported financially by Lembaga Pengelola Dana Pendidikan (LPDP). We would like to express our institution, Yogyakarta State University that has provided many experiences in term of research. Our deepest gratitude to students who participated in our experiment research and also who have contributed to the collection of our empirical data. Thanks to the teacher of senior high school who helped this research.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

REFERENCES

Abdurrahman, A., Setyaningsih, C.A., & Jalmo, T. (2019). Implementing multiple representation based worksheet to develop critical thinking skills. Journal of Turkish Science Education, 16(1), 138-155. http://dx.doi.org/10.12973/tused.10271a

Adams, R. J., & Khoo, S.-t. (1996). *Quest: The interactive test analysis system, Version 2.1*. Melbourne: ACER.

Anderson, L. W., Krathwohl, D. R., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., Raths, J., & Wittrock, M. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy*. New York: Longman Publishing.

Batlolona, J. R., Diatoro, M., Wartono., & Latifah, E. (2019). Creative Thinking Skills Students in Physics on Solid Material Elasticity. Journal of Turkish Science Education, 16(1), 48–61. (https://doi.org/10.12973/tused.10265a

Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational goals. *Cognitive domain*.

Boopathiraj, C., & Chellamani, K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International journal of social science & interdisciplinary research*, *2*(2), 189-193.

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*: ASCD.

Gall, M. D., Borg, R., & Gall, P. (1996). *Educational research: An instruction*. New York: White Plains: Longman.

Gronlund, N., & Linn, R. (1990). *Measurement and Evaluation in Teaching* 6th Ed. USA: Millan Publishing Company.

Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test… or testing to teach: exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, *26*(2), 307-329.

King, F., Goodson, L., & Rohani, F. (1998). *Higher order thinking skills*. Retrieved January, 31, 2011.

Mardapi, D. (2004). *Penyusunan tes hasil belajar*. Yogyakarta: PPS UNY.

Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes*. Jogjakarta: Mitra Cendekia.

Nguyễn, T. M. T., & Nguyễn, T. T. L. (2017). Influence of explicit higher-order thinking skills instruction on students' learning of linguistics. *Thinking Skills and Creativity*, 26, 113-127.

Ong, K. K. A., Hart, C. E., & Chen, P. K. (2016). *Promoting higher-order thinking through teacher questioning: a case study of a singapore science classroom*. New Waves, 19(1),1.

Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. (2019). The Development of an Instrument to Measure the Higher Order Thinking Skill in Physics. *European Journal of Educational Research*, *8*(3), 743-751. doi: 10.12973/eu-jer.8.3.743

Ramadhan, S., Sumiharsono, R., Mardapi, D., & Prasetyo, Z. K. (2020). The Quality of Test Instruments Constructed by Teachers in Bima Regency, Indonesia: Document Analysis. *International Journal of Instruction, 13*(2). doi: 10.29333/IJI.2020.13235A

Satria, K., & Uno, H. B. (2012). *Assesment Pembelajaran*. Jakarta: Bumi Aksara.

Tiruneh, D. T., De Cock, M., & Elen, J. (2017). Designing learning environments for critical thinking: Examining effective instructional approaches. *International Journal of Science and Mathematics Education*, 1-25.

Zohar, A., & Cohen, A. (2016). Large scale implementation of higher order thinking (HOT) in civic education: The interplay of policy, politics, pedagogical leadership and detailed pedagogical planning. *Thinking Skills and Creativity*, 21, 85-96.