P-ISSN 2746-9034 || E-ISSN 2774-5066
Indonesian Physics Education Research
Volume 03, No. 02, December 2022, pp. 87-97

DEVELOPMENT OF A HOT (HIGHER ORDER THINKING) TEST INSTRUMENT ON WORK AND ENERGY AT SMA NEGERI 1 SIBOLANGIT T.A 2022/2023

Jojor Sihol Marito Tambunan[1], Dewi Wulandari[1]
[1,2] Universitas Negeri Medan, Medan 20113, Indonesia
Email: jojortambunan297@gmail.com

Abstract

This study aims to develop a test instrument based on HOT (Higher Order Thinking) that meets the aspects of validity, reliability, level of difficulty, and discriminatory power. This type of research is development research with a 4-D model by Thiagarajan, including the define, design, develop, and disseminate stages. Small group tests were conducted on 10 students and large group tests were conducted on 30 students in class XI MIA 1 at SMA Negeri 1 Sibolangit. The results showed that the HOT (Higher Order Thinking) based test instrument on work and energy material consisted of 10 questions that were suitable for use, with expert validation test results obtained 80% valid items, small-scale test results obtained 83% valid items, test reliability of 0.8683 in the high correlation category, the difficulty level of the test in the medium category is 100%, the differentiability of the test in the good category is 80% and 20% is sufficient. The results of the large-scale test obtained test reliability of 0.79 with a high correlation category, the level of difficulty of the test were 70% moderate, 20% easy, and 10% difficult, and the distinguishing power of the test was 60% sufficient and 40% good.

Keywords: Research and development; Test Instruments; HOT (Higher Order Thinking)

Education is one of the important elements of human life. Through education, humans can maintain and improve their standard of living (Setiawati, 2017). Improving the quality of education is closely related to the demands of learning in the 21st century. According to Gates in Litna et al (2021), education is currently in the knowledge age with a very rapid increase in knowledge. This accelerated increase in knowledge is supported by digital media and technology called the Information Super Highway.

Education in Indonesia is currently implementing the latest curriculum, namely the revised 2013 curriculum. According to Amin et al (2018), this curriculum requires students to be able to predict, design, and predict. In line with this, the realm of Higher Order Thinking (HOT) includes the process of analyzing (C4), evaluating (C5), and creating (C6). Knowledge acquired through higher-order thinking processes is more easily transferred than simply memorizing. Through deep conceptual understanding, students can apply this knowledge to solve new problems in different situations. In addition, students are also prepared to have the skills needed in the 21st century, including critical, creative, problem-solving, collaboration, and communication. Therefore the assessment of student learning outcomes must include these aspects.

Assessment is not just collecting student score data, but also processing it to obtain an overview of the process and learning outcomes of students, and teachers must follow up on it for the benefit of learning. Therefore, the position of the learning outcomes assessment instrument is very strategic in teacher and school decision-making related to the achievement of student learning outcomes, which include higher-order thinking skills.

Learning will be meaningful if students are invited to think critically. Successful mastery of a concept will be obtained when students can think at a higher level. Where students not only remember and understand a concept but can also analyze and synthesize, evaluate, and create a concept well. Concepts that have been understood will stick in memory for a long time, for this reason, students need to have higher-order thinking skills or HOT.

Evaluations or assessments that are usually carried out by teachers only measure lower-order thinking skills (Lower Order Thinking), for example, memorizing formulas to solve questions without mastering the concept so that students' critical thinking skills cannot develop. The problems that occur in schools are questions that tend to test more aspects of memory that do not train students' higher-order thinking skills.

Based on the results of research conducted by Rahayu et al (2008), it is known that one of the factors causing students' thinking skills which tend to be low is the lack of training students in solving questions that measure higher-order thinking skills. The teacher's lack of ability to develop HOT instruments resulted in the unavailability of instruments specifically designed to train HOT, so it was necessary to develop HOT test instruments.

The development of students' higher-order thinking skills will result in better students' problem-solving skills, the level of students' confidence increases, and students' learning achievement in higher-order thinking increases (Budiman and Jailani, 2014).

Based on the results of initial observations through a questionnaire given to physics teachers at SMA Negeri 1 Sibolangit, information was obtained that in preparing test instruments for daily tests, assignments or exercises still tested knowledge, understanding, and application which only measured Lower Order Thinking (LOT) learners. Likewise, the results of the questionnaire given to students at SMA Negeri 1 Sibolangit said that learning physics was a difficult lesson to understand. This difficult learning makes students less interested in learning physics. Even though Indonesia has entered the era of the Industrial Revolution 4.0, where a student is required to be able to think critically. This can be realized by implementing a Higher Order Thinking (HOT) based test instrument in educational institutions.

Based on the description of the problems above, it is important to research the development of HOT (Higher Order Thinking) test instruments for Work and Energy.

METHODS

This type of research is research and development. Research and development methods (Research and Development) are used to produce certain products and test the effectiveness of these products. The product developed in this study is a test instrument in the form of a description based on HOT (Higher Order Thinking). The development design in this study uses the 4D model (four-D models) developed by Thiagarajan, consisting of 4 stages, namely: (1) Define; (2) Design; (3) Develop; (4) Disseminate. This research was conducted at SMA Negeri 1 Sibolangit and was conducted by students of class XI MIA 1, where the test instrument test consisted of a small group test phase consisting of 10 students and a large group test stage consisting of 30 students.

Define Stage

This stage aims to determine and define needs by analyzing the objectives and limitations of the material. In this development research, the material boundaries set are "Work and Energy". In this stage there are three steps of definition, namely:

1) Needs Analysis

   A needs analysis was carried out by direct observation to know the problems that exist in SMA Negeri 1 Sibolangit.

2) Task Analysis

   Task analysis is needed to determine the material to be used in the higher-order thinking test instrument.

3) Concept Analysis

   Concept analysis aims to identify concepts, arrange concepts systematically and relate existing concepts to create a test instrument to measure higher-order thinking in the matter of work and energy.

Design Stage

This stage aims to make a design for the product that has been set. In this stage, there are three design steps, namely:

a) Instrument Form Determination

   The form of instrument that will be used to measure high-level thinking in class X on work and energy is a written test in the form of an essay.

b) Arrangement of Grids

The making of the grid aims to determine the scope and is used for instructions for making questions. The test questions in this study are in the form of objective tests in the form of essays.

c) Instrument Design

This stage aims to create an initial framework for the assessment instrument to collect data. At this stage, questions are made, scoring guidelines, validation sheets, and instrument printing.

Develop Stage

This stage aims to make the design into a product and test the validity of the product repeatedly until the product is produced according to the specifications set. In this stage there are four development steps, namely:

1. Expert or Practitioner Validation

   This validation stage aims to improve the initial development of the instrument by experts or practitioners. The validation technique is carried out using a validation sheet as an assessment and input from experts or practitioners which will then be revised.

2. Trial Test

   Test trials were conducted to test the legibility of the developed higher-order thinking assessment instrument. The trial implementation of the test that was developed is small-scale and large-scale testing. The purpose of small-scale trials is to obtain the results of new product development on a small scale. On the other hand, large-scale trials aim to determine whether the product being developed meets its performance goals.

3. Item Analysis

   Item analysis was carried out to determine validity, reliability, item difficulty level, and discriminating power.

4. Student Response Test

   Analysis of student response tests aims to find out how students respond to the test that was developed, by giving a questionnaire containing questions.

Disseminate Stage

At this stage, instrument dissemination activities were carried out by providing assessment instruments to the school where the research was conducted at SMA Negeri 1 Sibolangit and the application of the instruments was carried out through a feasibility test in class implementation before being distributed outside the institution.

The data collection instrument in this study was a questionnaire sheet containing questionnaire instruments used to obtain validation data from the material validator. Student response questionnaire sheets contained questions filled in by students to find out student responses to the instruments developed by researchers, and there were grids. developed test instrument.

Data collection techniques aim to collect the data needed in research. Data collection was carried out during the process of preparing the assessment instruments as well as during the assessment process in class, including through:

(1) Testing the feasibility of the developed test instrument using expert judgment that will validate the contents of the instrument.
(2) Conducting trials of the test instruments developed for students of SMA Negeri 1 Sibolangit class X MIA 1.
(3) Documentation in the form of valuable data from the test results of test instruments conducted by students.

The data analysis technique used was expert validation analysis, test validity, reliability test, level of difficulty, and discriminatory power.

RESULT AND DISCUSSION

This research was conducted at SMA Negeri 1 Sibolangit, in class XI MIA 1 with 30 students. This type of research is research and development with a 4D development model consisting of (1) Define; (2) Design (design); (3) Develop (development); (4) Dissemination, the product developed is a HOT (High Order Thinking) Based Test Instrument on Work and Energy.

Expert Validation Results

After the initial design of the test instrument was completed, the researcher validated 3 validators, consisting of 2 (two) lecturers at Medan State University and 1 (one) a physics teacher at SMA Negeri 1 Sibolangit.

Table-1. Table of Expert Validation Results

Question Number	CVI	CVR	Category
1	1	1	Valid
2	1	1	Valid
3	1	1	Valid
4	1	1	Valid
5	1	1	Valid
6	1	1	Valid

7	1	1	Valid
8	1	1	Valid
9	1	1	Valid
10	1	1	Valid
11	1	1	Valid
12	1	1	Valid
13	0,54	1	Invalid
14	0,65	1	Invalid
15	0,40	1	Invalid

Assess the validity of an instrument, it is done by calculating the CVR (Content Validity Ratio), after which the CVI (Content Validity Index) is calculated by dividing the total number of CVR (Content Validity Ratio) by the number of questions developed. According to Lawshe CVR (Content Validity Ratio) and CVI (Content Validity Index), a value range of -1 and below 0 is included in the bad category, a value of 0 is included in the good category, and a value range greater than zero to 1 is included in the category very good.

Discussion

Small Group Test Results

Based on the results of the validity, reliability, difficulty level and discriminating power tests carried out in small groups, the results for the validity test were 10 valid questions and 2 invalid questions.

The items that are declared valid have a value $r\text{-}count > r\text{-}table$, where the value of $r\text{-}count$ in the valid questions is above 0.632. The valid questions consist of question number 1 with $r\text{-}count$ of 0.729; item number 2 with $r\text{-}count$ of 0.748; item number 3 with $r\text{-}count$ of 0.741; item number 5 with $r\text{-}count$ of 0.655; item number 6 with $r\text{-}count$ of 0.692; item number 7 with $r\text{-}count$ of 0.666; item number 8 with $r\text{-}count$ of 0.642; item number 9 with $rcount$ of 0.641; item number 10 with $rcount$ of 0.646; and item number 11 with $rcount$ of 0.655.

Meanwhile, the invalid questions are in question number 4 with $rcount$ of 0.038; and item number 12 with $rcount$ of 0.304.

Table-2. Test Instrument Validity in Small Group Trials

Question Number	r_{count}	r_{table}	Category
1	0.729	0.632	Valid
2	0.748	0.632	Valid
3	0.741	0.632	Valid

4	0.038	0.632	Invalid
5	0.655	0.632	Valid
6	0.692	0.632	Valid
7	0.666	0.632	Valid
8	0.642	0.632	Valid
9	0.641	0.632	Valid
10	0.646	0.632	Valid
11	0.655	0.632	Valid
12	0.304	0.632	Invalid

The reliability of the developed test instrument is at a value of 0.86833 with high-reliability criteria.

Tabel-3. Test Instrument Reliability in Small Group Trials.

Number of Questions	Reliability	Criteria
10	0.86833	High Reliability

On the results of the difficulty level test, it was found that there were no difficult and easy items, but there were 10 items in the medium category. In item number 1, the difficulty level is 0.30; item number 2 has a difficulty level of 0.33; item number 3 has a difficulty level of 0.48; item number 4 has a difficulty level of 0.55; item number 5 has a difficulty level of 0.58; item number 6 has a difficulty level of 0.48; item number 7 has a difficulty level of 0.48; item number 8 has a difficulty level of 0.53; item number 9 has a difficulty level of 0.40; and item number 10 has a difficulty level of 0.28.

Tabel-4. Difficulty Level of Test Instruments in Small Group Trials

Question Number	Difficulty Index	Category
1	0.30	Medium
2	0.33	Medium
3	0.48	Medium
4	0.55	Medium
5	0.58	Medium
6	0.48	Medium
7	0.48	Medium
8	0.53	Medium

9	0.40	Medium
10	0.28	Medium

In the results of the discriminating power test, there were 8 items in the good category, namely item number 1 having a discriminating power of 0.42; item number 2 has a different power of 0.42; item number 4 has a different power of 0.42; item number 5 has a different power of 0.50; item number 6 has a different power of 0.50; item number 7 has a different power of 0.42; item number 8 has a different power of 0.42; item number 9 has a different power of 0.50 and there are 2 items in the sufficient category, namely item number 3 with different power of 0.33 and item number 10 with different power of 0.33.

Tabel-5. Distinguishing Power of Test Instruments Small Group Trials

Question Number	Distinguishing Power	Category
1	0.42	Good
2	0.42	Good
3	0.33	Sufficiently
4	0.42	Good
5	0.50	Good
6	0.50	Good
7	0.42	Good
8	0.42	Good
9	0.50	Good
10	0.33	Sufficiently

Results of Large-Scale Trials

After carrying out a small group test with a total of 10 students, then a large group test was carried out with a total of 30 students with a total of 10 items in the test instrument. From the large group test, the results for the reliability of the developed test instrument were at a value of 0.79 with high-reliability criteria.

Tabel-6. Test Instrument Reliability in Large Group Trials

Reliability	Criteria
0.79	High Reliability

On the difficulty level test results, it was found that there were 2 items in the easy category, namely item number 3 with a difficulty level of 0.83 and item number 6 with a difficulty level of 0.82. There are 7 items in the moderate category, namely item number 1 with a difficulty level of 0.60; item number 2 with a difficulty level of 0.33; item number 4 with a difficulty level of 0.32; item number 5 with a difficulty level of 0.58; item number 7 with a difficulty level of 0.75; item number 8 with a difficulty level of 0.69; and item number 10 with a difficulty level of 0.53. There is 1 item with a difficult category, namely item number 9 with a difficulty level of 0.25.

Tabel-7. Difficulty Level of Test Instruments in Large Group Trials

Question Number	Difficulty Level	Criteria
1	0.60	Medium
2	0.33	Medium
3	0.83	Easy
4	0.32	Medium
5	0.58	Medium
6	0.82	Easy
7	0.75	Medium
8	0.69	Medium
9	0.25	Difficult
10	0.53	Medium

In the results of the discriminating power test, there were 4 items in the good category, namely item number 1 having a discriminating power of 0.42; item number 2 has a different power of 0.42; item number 5 has a different power of 0.42; item number 10 has a different power of 0.50 and there are 6 items in the sufficient category, namely item number 3 with a different power of 0.33; item number 4 with a different power of 0.33; item number 6 has a different power of 0.34; item number 7 has a different power of 0.38; item number 8 has a different power of 0.28; and item number 9 has a different power of 0.38.

Tabel-8. Distinguishing Power of Test Instruments Large Group Trials

Question Number	Distinguishing Power	Criteria
1	0.47	Good
2	0.53	Good
3	0.25	Sufficiently
4	0.28	Sufficiently

5	0.50	Good
6	0.34	Sufficiently
7	0.38	Sufficiently
8	0.28	Sufficiently
9	0.38	Sufficiently
10	0.47	Good

CONCLUSION

Based on the results and discussion of the research that has been described previously, it is concluded that: (1) The HOT (Higher Order Thinking) based test instrument developed on Work and Energy in class XI MIA 1 SMA Negeri 1 Sibolangit consists of 15 questions in essay form. From the results of the field test, it was found that there were 10 items in the valid category, where these items had a value of $rcount > 0.632$. (2) In the feasibility test of the developed test instrument, reliability, difficulty level, and discriminating power tests were carried out. Based on the results of trials conducted in class XI MIA with 30 students participating, it was found that the developed test instrument had a high reliability of 0.79. In the difficulty level test, there are 2 items in the easy category, 7 items in the medium category, and 1 item in the difficult category. In the different power tests, 4 items were obtained in the good category and 6 items in the sufficient category. So it can be concluded that the test instrument developed is in the feasible category. (3) The results of the student response questionnaire show that the test instrument was developed very well. This can be seen from the achievement score obtained above 80%. This explains that the developed test instrument is feasible to be used as a measuring tool to measure the cognitive level of students.

REFERENCES

Amin, D. I., Sutrisno, S., & Sigit, D. (2018). Instrumen asesmen pemahaman konseptual berorientasi higher order thinking skills keterampilan proses dan sikap terhadap sains pada bahan kajian hidrokarbon dan minyak bumi. Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan, 3(9), 1142-1146.

Brookhart, S.M. (2010). How to Assess Higher Order Thinking Skills in Your Classroom. Alexandria: ASCD.

Budiman, A., & Jailani, J. (2014). Pengembangan instrumen asesmen higher order thinking skill (HOTS) pada mata pelajaran matematika SMP kelas VIII semester 1. Jurnal Riset Pendidikan Matematika, 1(2), 139-151.

Lawshe, C.H. (1975). A Quantitative Approach to Content Validity. Personnel Psychology, 28: 563-575.

Litna, K.O., Mertasari, N.M.S. & Sudirtha, G. (2021). Pengembangan Instrumen Tes Higher Order Thinking Skills (HOTS) Matematika SMA Kelas X. Jurnal Penelitian dan Evaluasi Pendidikan Indonesia, 11 (1): 10-20.

Rahayu, T., Purwoko, P., & Zulkardi, Z. (2008). Pengembangan Instrumen Penilaian dalam Pendidikan Matematika Realistik Indonesia (PMRI) di SMPN 17 Palembang. Jurnal Pendidikan Matematika, 2(2) : 17-33.

Setiawati, N.A. (2017). Pendidikan Karakter sebagai Pilar Pembentukan Karakter Bangsa. Prosiding Seminar Nasional Tahunan Fakultas Ilmu Sosial Universitas Negeri Medan. 1(1): 348-352.