

Developing Lup Instrument Test to Measure Higher Order Thinking Skills (HOTS) Bloomian for Senior High School Students

Lilis Erfianti¹, Edi Istiyono², Heru Kuswanto³

ARTICLE INFO

Article History:

Received 19.12.2018

Received in revised form

15.02.2019 Tarih girmek

için burayı tıklayın.

Available online 01.07.2019

ABSTRACT

The study aims to developing lup instrument test for measuring higher order thinking skills of students in physics learning. The 4D stage of the model consists of define, planning, design, and disseminate stages. The assessment instrument was developed based on HOTS indicators include the ability to analyze (C4), evaluate (C5), and creat (C6). The test form is a reasoning multiple choice. Subject of this research were all student of XI MIA 1 and XI MIA 2 on 2nd semester at SMA Negeri 1 Depok in academic year of 2017/2018. Result of this research are: 1) the higher order thinking skills test instrument proved valid and worthy of use in the learning process. Based on expert validation, the test is valid with Aiken V index value of 0.89, the average value and standard deviation of INFIT MNSQ is 1.01 ± 0.09 , and has a good degree of difficulty with a range of difficulty levels between - 1.29 to 1.35. The higher order thinking skills test can also be used to measure students higher order thinking skills in physics learning with very low to excellent categories.

©IJERE. All rights reserved

Keywords:

Assessment, test, reasoning multiple choice, higher order thinking Bloomian, lup.

INTRODUCTION

Physics is a branch of science regarding understanding concepts and their application in problem solving. To achieve this goal conceptual change is needed by modifying the students' initial knowledge to match the actual scientific knowledge (Atasoy, 2009). In general, when students entering the class, they have wrong knowledge that built by themselves as a result of interaction with nature (Suparno, 2013). Conceptual change requires learning that allows students to develop new concepts and improve previous ways of thinking (Atasoy, 2009; Arends, 2012; Frank & Scherr, 2012; Treagust & Duit, 2008). Therefore, students need facilities to keep knowledge in depth for a long period of time so that knowledge can be activated when needed.

The concept of learning physics has not been well conveyed so far (Barniol & Zavala, 2014). The failure of students in solving problems is caused more by failure in activating relevant knowledge that is already owned, not because they do not have the right knowledge related to the problem (Hammer, 2000). So, they need direction to activate knowledge quickly and accurately when solving physics problems. Assessment is the heart of teaching and learning (Brooks, 2002). Instruments of assessment can be form on written test, oral, or observation sheets. Test can use to measure or find out learning outcomes of students after or before learning process is done. Multiple choice tests are one form of test that is often used, because its practical and easy in scoring. However, the usual form of multiple choice tests still has some weaknesses. Therefore, it is necessary to make excuses for students to choose these answers. The reasoning multiple choice test is not only able to measure student learning outcomes, but can reduce students to guess answers so that the ability of students can really be measured properly. In addition, a concept is needed understanding and thinking ability of students to answer correctly questions with reasoned multiple choice forms. High-level thinking skills are an important aspect of the learning process. High-level thinking skills are considered important as educational goals (Zohar & Dori, 2008). The development and improvement of HOTS students has become the main goal of education (Yen & Halili, 2015). The reasoned multiple choice test developed refers to Bloomian's higher order thinking skills (HOTS).

Assessment is a statement of a number of facts to explain the characteristics of a person or something (Griffin, Care & McGraw, 2012). Assessment is not only limited to students, but also assessment of educators, learning methods and school administration. The most effective test is used to find out and identify the level of understanding of students (Westwood, 2008). By using tests, learning outcomes and the

¹ Corresponding e-mail: erfianlilis@gmail.com, <https://orcid.org/0000-0003-4415-0054>

² <https://orcid.org/0000-0001-6034-142X>

³ <https://orcid.org/0000-0002-2693-8078>
Yogyakarta State University^{1,2,3}

level of understanding of students can be known as part of information that can help improve the learning process

Assessment instruments must meet the requirements of substance (content), construction and language and have evidence of empirical validity and produce scores that can be compared between schools, between regions and between nationalities (Mendiknas, 2007). In addition, standard assessment instruments must be valid, reliable, and free from bias elements (Mardapi, Kumaidi & Kartowarigan, 2011). Objective test instruments can be classified into four types. Correct test, matching test, short description test, and multiple choice test. Multiple choice tests are most commonly used for tests (Jeong, 2012; Nikou & Economides, 2016). Multiple choice tests consist of a description of knowledge that must be completed by choosing one of several possible answers provided (Jeong, 2012). A two-level selection test or also called a multiple choice test reasoned has three roles: (a) Diagnostic tool, which is to identify students' initial knowledge, misconceptions and other concepts (Won, Yoon & Treagust, 2014), (B) Instructional Tools, and (C) Learning Devices, namely to help learners improve misconceptions and improve their understanding (Chua & Don, 2013). Judging from the reasons for this multiple choice, it is able to correct the deficiencies that exist in multiple-choice one level, which can minimize the answers of test participants who answer correctly by guessing.

The assessment process used has not been able to measure students' thinking skills. The form of instruments commonly used by teachers in assessment are test instruments which include multiple choice questions or descriptions. The questions used by the teacher are only limited to measuring students' knowledge. The indicators used in the question grid are still at the lower order level of thinking skills (LOTS), namely C1, C2, and C3. So that the understanding and ability of high-thinking students is not really well measured. Therefore, it is necessary to develop test instruments that are appropriate for students at this time.

Students' thinking skills are grouped into two categories, namely (1) low-level thinking (LOTS) consisting of knowledge (C1), understanding (C2), and application (C3); and (2) high-level thinking (HOTS) consisting of analysis (C4), evaluation (C5), and creating (C6) (Schraw, Gregory & Robinson, 2011). The test of the learning outcomes of high school physics subjects in the form of multiple choices is limited to measuring low-level thinking skills (LOTS) and has not been able to measure the ability of higher-order thinking (HOTS) of students (Istiyono, Mardapi & Suparno, 2014)). Therefore a test is needed to be able to measure HOTS students. The form of tests that are commonly used are multiple choice questions, while the multiple choice questions are rarely able to measure HOTS students.

One form of test developed in this study is reasoning multiple choice. In addition, so that the students' high-thinking skills can be measured properly, the question indicators use HOTS Bloomian aspects, namely at levels C4, C5, and C6. In this study researchers limit the study population, where the study population used in this study were all students in two classes of XI MIA.

METHOD

This research is a development research better known as 4D. The 4D stage of the model consists of define, planning, design, and disseminate stages. This research was conducted at SMAN 1 Depok, Jl. Babarsari, Kel. Caturtunggal, Kec. Depok, Kab. Sleman, Caturtunggal, Sleman, Sleman Regency, Yogyakarta Special Region 55281, in the even semester of 2017/2018 Academic Year. Tests to determine the feasibility of the questions and the level of difficulty of the items were conducted on 18 students of the XII MIA SMA N 1 Depok class. Reasonable multiple choice test consisting of 10 items that pass the validation process. The results of the trial were analyzed using SPSS to find descriptive statistics while the analysis of the items was analyzed using the QUEST program to determine the students' ability profile. For modeling and implementation classes, the score of the test results is categorized using a scale of five. Data retrieval of learning outcomes was conducted for 64 students from XI MIA 1 and XI MIA 3 to find out the mapping of high-level thinking skills in physics learning at the school.

Material

The instrument used for collecting data consists of grids arranging question, two-tier multiple-choice, answer key and assessment rubrics, and instrument assessment validation sheets developed. The result of the test were analyzed by estimation result, goodness of fit, and item difficulty level.

Data Analyses

Qualitative analysis of data in the form of the value of the validity of the contents of the standard cognitive test instrument is based on the criteria of Rahayu (2012) as in Table 1 below .

Table 1. Interpretation Criteria for Validity of Test Items

Validity	Interpretaton of Item Validity
0.80-1.00	Very High
0.60-0.80	High
0.40-0.60	Medium
0.20-0.60	Low
0.00-0.20	Very low

Source : Rahayu (2012)

In the response theory approach, one parameter logistic model, the quality of the questions can be grouped as follows:

- a. Good, if the item matches the model and the difficulty level is $-2.00 \leq b_i \leq 2.00$
- b. Good enough, if the item matches the model and the difficulty level parameter ≤ -2.00 or ≥ 2.00 .
- c. Not good, if the item doesn't match the model

The data analysis was counted using QUEST and SPSS. Item estimation, goodness of fit, and item difficulty can describe using QUEST. An item or testee is suitable to the model in the range of INFIT MNSQ of 0.77 to 1.30. SPSS used to descrbe ability of students from the result of the test score. It also know the frequency and percentage of students who are in the category low to excellent.

FINDINGS

The process of developing HOTS Test Instruments on Lup

The test instrument developed in the form of reasoning multiple choice consists of the choice of answers and reasons for choosing the answer with five choices of answers and reasons. The process of developing test instruments is:

- 1. Product Design

This stage is carried out in accordance with the test development stages. The initial step in developing the test is the preparation of test instrument specifications and assembly questions

- 2. Guided Instrument

The preparation of the instrument grid was adjusted to the KD achievement indicator, the 2013 HOTS curriculum indicator. The form of the test instrument developed was an objective reasoning multiple choice with five answer choices. The length of the instrument developed is 10 questions with a time allocation of 30 minutes. After preparing the specification of the test instrument, it is done by assembling the questions. Each item was written based on the formulation of the indicators that had been arranged in the grid, taking into account the rules of writing multiple choice questions with good reasoning. The queston is shown in table 2.

Table 2. Grid of Test Instrument

HOTS Indicators	Sub aspects of HOTS	Question Indicators	Questions
Analyze	Finding	Students can	Imam repairing his watch that was

	<p>find Lup function</p>	<p>damaged. He uses several tools, on of them is Lup. Lup are used to make it easier to see the watch component to ber repaired. The statement above shown that....</p> <ol style="list-style-type: none"> The lup or magnifying glass serves to enlarge the object being observed. lup or magnifying glass serves to enlarge the shadow of the observed object lup or magnifying glass serves to increase the distance of the object being observed Lup or magnifying glass serves to increase the distance of the observed shadow Lup or magnifying glass serves to increase the height of the object being observed <p>Reason:</p> <ol style="list-style-type: none"> The object is located at the Focus of lup that composed by concave lens The object is located on the radius of the lup that composed of concave lens Objects are located in the stage I of lup which is composed of a convex lens Objects located in stage II, which are composed of convex lenses The object is located in the stage III that composed of a convex lens
<p>Finding</p>	<p>Students can find Lup function</p>	<p>Fita is an anthropologist. He found ancient artifacts, but he had difficulty analyzing the contents of the artifact even though he had tried to read it as close as possible. To see clearly, Fita should ...</p> <ol style="list-style-type: none"> minimize the distance of the artifact to be more visible to the eye use a lup to enlarge the shadow use a lup to enlarge objects. use a lup in the form of loops to enlarge the distance of objects using a lup in the form of lup to enlarge height <p>Reason:</p> <ol style="list-style-type: none"> The smaller the distance of the object, the more visible the shadow will be Lup are needed by the eye, because the eye cannot see objects that are smaller than the point near the eye. Lup are needed by the eye, because the eye cannot see objects that are farther away from the eye near the eye. Lup are needed by the eye, because the eye cannot capture shadows that are smaller than the point near the eye. Lup are needed by the eye, because the eye cannot capture shadows that are larger than the point near the eye

3. Validation

Validation of the questions was carried out to determine the extent of the feasibility of the test instrument products developed based on the data obtained by experts. Validation results are then analyzed and used as consideration in revising the initial product.

Assessment of higher order thinking skills test instruments is reviewed based on subject, construction, and language. Each item is assessed based on three aspects of assessment. Based on the response and assessment of the two validators, the V Aiken index value of the 10 items was in the range 0.88-0.91, with the overall V aiken index can be seen in Table 3.

Table 3. Validity of item

Item	Aiken V	Category
1	0.90	Very high
2	0.91	Very high
3	0.88	Very high
4	0.90	Very high
5	0.90	Very high
6	0.91	Very high
7	0.89	Very high
8	0.88	Very high
9	0.90	Very high
10	0.90	Very high
Rata-rata	0.89	Very high

Based on the data above the value of the index V aiken the overall problem is at 0.89-0.91, so that the whole item that has been developed is valid according to the expert judgment and can be used in the next test phase after going through the revision stage. As for some items, namely items number 3, 7, and 8 were revised according to the advice of experts. The revised aspect is regarding the suitability of the test with KI, KD, and question indicators

4. Empiric Validation

After going through expert validation, the test instrument was empirically tested. Higher order thinking skills test contains 10 questions that will be used in the empirical trial stage. An empirical trial was carried out on high school students of class XII in SMA 1 Depok. The number of respondents in the trial was limited to 22 students.

a. Estimation Result

Based on the calculation of Quest, it is obtained the estimation of higher order test items thinking skills. Item estimates are shown in Table. 4

Tabel 4. Estimated Items and Testies

No	Item Estmation	Testies Estimation	
1	0.04±0.79	0.01±0.26	Average and Standard Deviation
2	1.01±0.09	0.99±0.44	Average and Standard Deviation INFIT MNSQ
3	0.08±0.51	0.08±0.51	Average and Standard Deviation INFIT t
4	0.13±0.38	0.12±0.73	Average and Standard Deviation OUTFIT t

5	0	0	Item atau score testies
6	0	0	Item atau testi skor <i>perfect</i>

b. Goodness of Fit

Based on the results of the empirical trials analyzed using Quest, the results of goodness of fit testing for the test in each item with the Quest program were obtained. fit test tests the whole item developed based on the mean of INFIT Mean of Square (Mean INFIT MNSQ) and standard deviation or observes the average value of INFIT-t.

The item acceptance limit uses MNSQ infit between 0.77 to 1.30. Distribution of test items according to Quest's output is more clearly shown in table 5 below.

Table 5. Goodness of fit

ITEM	MNSQ INFIT	
	0.77-1.00	1.00-1.30
Item 1		√
Item 2		√
Item 3		√
Item 4		√
Item 5		√
Item 6		√
Item 7		√
Item 8		√
Item 9	√	
Item 10	√	

These results conclude if the items developed are fit with the Partial Credit Model (PCM) model and valid based on the empirical test with the distribution of Goodness of Fit based on INFIT MNSQ 10 items about Higher Order Thinking Skills being between the range 0.77-1.30.

c. Item Difficulty Level

The level of difficulty per item can be seen from the output it is by showing the value of item difficulty level.

Table 6. Item Difficulty Level

No.	Item difficulty level
1	-0.98
2	-0.96
3	0.79
4	0.16
5	-0.20
6	1.35
7	0.72
8	-1.29
9	0.27
10	0.16

Based on the data from Quest analysis, the item with the highest level of difficulty is item 6, which is 1.35. Whereas the problem with the lowest level of difficulty is the 8th item with value of -1.29.

Result of Higher Order Thinking Skills

1. Experimental Class

The results of students' high-level thinking skills are obtained from the results of the test analysis. The following are the results of the test analysis using SPSS in the form of a statistical test description.

Table 7. Descriptive statistics for experimental class

	N	Minimum	Maximum	Std.		
				Mean	Deviation	Variance
Result	32	60.00	82.50	67.03	6.23	38.8

Based on the table, it show minimum and maximum scores. At this experimental class, a minimum score is 60 and a maximum score is 82.5, with a standard deviation of 6.23. In addition, the average score generated by 32 students was 67.03

Table 8. Frequency of Experimental Class test result

Test Score	Frequency	Category	Persentase (%)
60.00	5	Very Low	15.6
62.50	7	Very Low	21.9
65.00	5	Very Low	15.6
67.50	6	Low	18.8
70.00	2	Medium	6.3
72.50	2	Medium	6.3
75.00	2	High	6.3
77.50	1	High	3.1
82.50	2	Excellent	6.3

In the table can be known the ability of students from the results of the test scores obtained. Besides that, it is also known that the frequency and percentage of students who are in the category are deficient to very good. In this modeling stage all categories are in the analysis results.

2. Implementation Class

The results of students' high-level thinking skills are obtained from the results of the test analysis. The following are the results of the test analysis using SPSS in the form of a statistical test description.

Table 9. Descriptive statistics for implementation class

		Minimum	Maximum	Mean	Std. Deviation	Variance
Value	32	60.00	87.50	68.3594	7.05992	49.842

Based on the table, it is known minimum and maximum scores. At this implementation class, the minimum score is 60 and the maximum score is 87.5, with a standard deviation of 7.059. In addition, the average value generated by 30 students is 68.35.

Table 10. Frequency of Implementasi Class test result

No	Skor Tes	Frekuensi	Kategori	Persentase (%)
1	62.50	6	Very low	18.8
2	65.00	6	Very low	18.8
3	67.50	6	Low	18.8
4	70.00	3	Low	9.4
5	72.50	5	Medium	15.5
6	75.00	2	Medium	6.3
7	77.50	1	High	3.1
8	82.50	1	Excellent	3.1
9	85.00	1	Excellent	3.1
10	87.50	1	Excellent	3.1

In the table can be known the ability of students from the results of the test scores obtained. Besides that, it is also known that the frequency and percentage of students who are in the category are very low to excellent.

RESULT, DISCUSSION AND SUGGESTION

The test instrument for measuring higher order thinking skills students in high school consisted of 10 reasoned multiple choice question. Where there is a choice of answers and the reasons for choosing the answer. In the preparation of the test instruments, several step were taken there are preparation of test spesification, writing test question, reviewing test question, testing, analyzing, revision, assembling and carrying out test (Mardapi, 2008: 88). Validation is grouped into three categories, namely content, construct, and criteria validity (Azwar, 2016: 40). The result of validation of the test instrument showed the test were arranged valid with an average 0.89 with very high categories. The quality of the items in the IRT is determined by the suitability of the items with the Rasch model and item difficulty index. The results of the study show that the goodness of fit value of all items is between two boundary lines, namely 0.79-1.13. Selain kecocokannya dengan model, hal lain yaitu tingkat kesukaran butir. The item is said to be good if the index of difficulty is more than -2.0 or less than 2.0. This refers to the theory developed by Hambleton & Swaminathan (1985). Based on the calculation of the magnitude of the difficulty level of 10 items about Higher Order Thinking Skills developed in the range -1.29 to 1.35.

Higher order thinking skills of students can be seen from the result of the test score. Tests are carried out n two classes namely the experimental class and the implementation class. Each consists of 32 students. The test instrument used is the same, in the expermental class the researcher teaches students directly about material to be evaluated using the instrument. Whereas in the implementation class , teacher will teach students with the same material as the experimental class. From the result obtained from 64 students there were 9 students with higher order thinking skills (HOTS), 11 students in medium order thinking skills (MOTS) and 44 students in low order thinking skills (LOTS).

Based on the results of learning innovations and analysis results, the lup instrument test to measure higher order thinking skills bloomian consisting of 10 questions proved valid and feasible to be used in the learning process. Based on expert validation, the test was declared valid. Higher order thinking skills tests can also be used to measure students' high-level thinking skills in physics learning with very low to excellent categories.

REFERENCES

- Arends, R.I (2012). *Learning to teach: 9th edition*. New York: McGraw-Hill.
- Atasoy, B. (2009). The effect of a conceptual change approach on understanding of students' chemical equilibrium concept. *Journal Research and Science & Technological Education*, 27: 267–282.
<https://doi.org/10.1080/02635140903162587>
- Azwar, S. (2016). *Metode penelitian*. Yogyakarta: Pustaka Pelajar
- Barniol, Pablo., & Zavala, Genaro (2014). *Test of understanding of vectors: A reliable multiple choice vector concept test*. *Physics Education Research*, 10,12-13 <http://dx.doi.org/10.1103/PhysRevSTPER.10.010121>
- Brooks, V. (2002). *Assessment in secondary school: The new teacher's Guide to monitoring ,assessment, recording, reporting and accountability*. Philadelphia: Open University Press
- Chua, Y. P., & Don, Z. M. (2013). Effects of computer-based educational achievement test on test performance and test takers' motivation. *Computers in Human Behavior*,29(5),1889-1895.
<https://doi.org/10.1016/j.chb.2013.03.008>
- Frank, W., & Scherr, R. (2012). Interactional processes for stabilizing conceptual coherences in physics. *Physical Review Special Topic-Physics Education Research*, 8.
<https://doi.org/10.1103/PhysRevSTPER.8.020101>
- Griffin, P., Care, E., & McGaw, B. (2012). *The changing role of education and school*. Springer.
http://doi.org/10.1007/978-94-007-2324-5_1
- Hambleton, R. K., & Swaminathan, H. (1985). Item response theory, principles and criteria: An example from wine ratings and prices. *Wine Economics and Policy*, 3(1) -9
- Hammer, D. (2000). Students resource for learning introductory physics. *American Journal of Physics*, *Physics Education Research Supplement*, 68 (S1), S52–S59
- Istiyono, E., Mardapi, D., & Suparno (2014). *Pengembangan Instrumen Untuk Mengukur Kemampuan Berfikir Tingkat Tinggi Fisika (PsyTHOTS) Peserta Didik SMA*. Jurnal Penelitian dan Evaluasi Pendidikan
- Jeong, H. (2012). A comparative study of score on CB and PBT. *Behaviour and information Technology*
<https://doi.org/10.1080/0144929X.2012.710647>
- Mardapi, D. (2008). *Teknik Penyusunan Instrumen dann Nontes*. Yogyakarta: Mitra Cendekia Offset.
- Mardapi, D., Kumaidi. & Kartowagiran, B. (2011). Pengembangan Instrumen Pengukur Hasil Belajar Nirbias dan Terskala Baru. *Jurnal Penelitian dan Evaluasi Pendidikan* 15(2): 326-341
- Mendiknas (2007). Permendiknas Ri No.20 Tentang Standar Penilaian Pendidikan
- Nikou, S. A., & Economides, A. A. (2016.) The impact of paper based, computer based and mobile based self assessment on students science motivation and achievement. *Computers in Human Behaviour*, 55, 1241-1248 <https://doi.org/10.1016/j.chb.2015.09.025>
- Rahayu, D & U. Azizah. (2012). Pengembangan Instrumen Penilaian Kognitif Berbasis Komputer dengan Kombinasi Permainan "Who Wants To Be A Chemist" pada Materi Pokok Struktur Atom untuk Kelas X SMA RSBI. *Prosiding Seminar Nasional Kimia UNESA*. Hal. 41-50
- Schraw , Gregory, & Robinson, D.H. (2011). *Assessment of higher order thinking skills*. New York: Information Age Publishing, Inc ISBN: ISBN-978-1-6173-5506-6
- Suparno, Paul., SJ. (2013). *Metodologi Pembelajaran Fisika Konstruktivistik dan Menyenangkan*. Yogyakarta: Universitas Sanata Darma
- Treagust, D. F., & Duit, R. (2008). *Conceptual change: A discussion of theoretical, methodological and practical challenges for science education*. Handbook on Conceptual Change. Perth: In Press
<https://doi.org/10.1007/s11422-008-9090-4>
- Westwood, P. (2008). *What teachers need to know about teaching methods*. Victoria: Acer Press

- Won,M., Yoon, H., & Treagust, D.F.(2014). Students' learningstrategies with multiple representations: Explanations of the human breathing mechanism. *Sci. Educ.*,98(5), 840–866
<https://doi.org/10.1002/sce.21128>
- Yen, T.S., & Halili, S.H. (2015). Effective teaching of higher-order thinking (hot) in education. *The Online Journal of Distance Education and e-Learning*, 3(2), 41-47
- Zohar, A., & Dori, Y. J., (2008). Higher order thinking skills and low achieving students: Are they mutually exclusive? *The Journal Of The Learning Sciences*, 12(2), 145-181.