

Development of Objective Test for High-Level Knowledge of High School Physics Materials on Temperature and Heat

Rizki Nursa'adah^{1,*}, Sahyar², Wawan Bunawan³

^{1,2,3}Physics Education Study Program of Postgraduate School of Universitas Negeri Medan, Indonesia 20221

*Corresponding author. Email: rizkinursaadah01@gmail.com

ABSTRACT

This study aims to develop an objective test of high-level knowledge of the subject of Temperature and Heat in high school that meets good test qualification standards in terms of validity, reliability, level of difficulty, discriminatory power and effectiveness of distractors. This type of research is development research, using the ADDIE research development method with the stages of analysis, design, development, and evaluation. In the analysis stage, the results obtained are minimal variations in the assessment and students are still lacking in obtaining physics questions for high-level knowledge of the temperature and heat material. The material about physics developed includes temperature, heat, expansion, Black's Principle, heat transfer, changes in the state of matter, ideal gases and thermodynamics. In the design phase, 50 items were obtained based on high-level knowledge analysis. The development stage obtained 38 items in the very valid category with a percentage of 92.4% based on content validation by 3 experts, both in terms of material, construct, and language. The results of the implementation and evaluation stages show that of the 50 items developed, 39 items have met the criteria for good test quality, based on the validity, reliability, discriminating power, level of difficulty, and effectiveness of distractors in small group trials and group trials big.

Keywords: High Level Knowledge Ability Test, Temperature and Heat, Validity, Reliability.

1. INTRODUCTION

During this 21st century, students are asked to be able to adapt to the increasing demands of the times. Students must produce higher-order thinking skills so that they are able to coordinate problem-solving skills by evaluating, analyze and creating or creating an idea, estimate or conception that can be strengthened by intelligent people with high-level skills. Because the ability to think at a higher level is one measure of the strength of one's intelligence. Where education plays an important role in the 2013 curriculum which is very significant in shaping human resources[1].

Regulation of the Minister of Education and Culture Number 54 of 2013 requires the competence of graduates to include attitudes, knowledge and skills. Meanwhile, the implementation of the 2013 curriculum has implications for the student's ability assessment model. Ability assessment by educators is carried out to see the way, progress, development of student

competency achievement in accordance with the expected potential and abilities on an on going basis. Assessment is a very important factor and cannot be separated from the act of learning. Assessments are completed in an orderly and structured manner. However, in measuring learning achievement, there are not always available instruments that match the learning objectives[2].

The test is a tool that is structured, objective and aims to obtain facts about many people or things that are desired in the right way[3]. According to Ashadarini, et al (2018), students only memorize formulas without mastering the concepts so that they experience difficulties in education. Based on this research, it can be concluded that students need education that can improve their conceptual skills[4].

Looking at the reality on the ground, there are still many educations that are trapped in low-level thinking skills, meanwhile education with higher-level thinking

skills is needed by students to compete and experience challenges in the future. This is in line with the research conducted by Karplus (Mirawati, 2011) that there are still many high school students who face difficulties in solving problems that require abstract thinking efficiently. Not only that, considering the fact that a student's physics achievement is measured on the reasoning aspect, Indonesia is ranked 40 of 42 countries[3]. To see the success of students in understanding concepts, it can be seen how capable students are in large-level thinking skills, where students can not only remember and master a concept, but also students can analyze, evaluate, and create a concept well. Concepts that have been understood well, of course, will stick in students' memories. So it is very important that students are guided towards higher-order thinking skills[5].

Some of the questions only demand students' memory abilities and solve mathematical formality problems in solving them, so that these kinds of questions are not effective and efficient when they become a tool for assessing students' level of understanding and higher-order thinking skills. The problems that exist should be analytical, not just practicing memorizing formulas. Basically, if you have mastered the concept, then the formula will also be obtained, because the formula in physics is the shrinkage of a concept [4]. Therefore, an objective test of students' high-level knowledge skills needs to be carried out to assess or measure the level of students' knowledge of concepts and theories of physics.

This research on developing higher order thinking tests has been carried out on various materials and sub-materials of physics. Among others are Validation and Structural Analysis of the Kinematics Concept Test [5], Development of Higher Order Physics (PsyHOTS) Thinking Ability Test for High School Students [6], Development of a Four-Stage Misconception Diagnostic Test on Kinematics (Pujayanto, et al, 2018), Development of Cognitive Learning Outcomes Test Instruments for Physics Subjects on the Subject of Momentum and Impulse [7], Development of Multiple Choice High Order Thinking Test Instruments in E-Learning-Based Physics Learning in SMA [8] dan Measuring Critical Thinking in Physics: Development and Validation of a Critical Thinking Test in Electricity and Magnetism [9]. Supported by many important studies, there are differences between previous research and research that will be carried out by researchers including the material, level of education studied, research methods and others. No one has researched the development of an objective test of high-level knowledge of high school physics on temperature and heat material.

The main way to evaluate a student's high level of knowledge in physics knowledge is by multiple choice trials [5]. Li & Singh (2016) said that the multiple-choice trial gave birth to many advantages that the other

test classes did not have. The advantages of multiple-choice trials are, (1) the arrangement is more representative in terms of covering and representing the number of subjects (2) it requires the examiner to behave objectively, both in correcting answers, deciding scores, and also as in deciding the final score (3) of things time effectiveness, correcting multiple choice is much faster (4) examination of the test can be done by anyone; (5) the test is much easier to analyze, both in terms of the level of difficulty, discriminatory power and reliability[4]. So multiple-choice trials are suitable for use in large-participant action situations with short processing times and extensive material[5].

The solution to increase the effectiveness of the test can be done by developing good trials and according to the needs of the times so that the quality of the trials will increase and the quality of education will be higher. Literature studies, preliminary product investigations, expert validation, small group trials, instrument validity and reliability tests, large group trials, data analysis and data processing to the formation of high-level knowledge ability test instruments are the movements carried out in this research. Supported by the above background, a research will be conducted with the title "Development of an Objective Test of High-Level Knowledge Ability in High School Physics on Temperature and Heat Materials".

2. METHOD

This research is a type of development research (Research & Development). According to Sugiyono (2018) the development research method is a research method used to produce certain products, and test the effectiveness of these products.

The population in this study were students in two high schools, namely SMA Negeri 2 Perbaungan and SMA Setia Budi Abadi Perbaungan. Sampling was carried out non-randomly, that is, one class was used as a sample for the small group scale test and four other classes were used as samples for the wide scale test.

The concept used is the development of Analysis, Design, Development, Implement and Evaluation (ADDIE) by Robert Maribe Branc, namely, analyze, design, develop, implement, and evaluate.

3. RESULT AND DISCUSSION

3.1. Result

This research was conducted to determine the quality of the items of the High-Level Knowledge Ability Test on Temperature and Heat material in SMA seen from qualitative analysis covering aspects of material, construction and language, as well as from the results of quantitative analysis including validation, reliability, level of difficulty, discriminating power and effectiveness of distractors. This research is a research and development (Research & Development) using the

stages of the ADDIE model which includes analyze, design, development, implementation and evaluation. The product of this development research is a matter of High Level Knowledge ability on the material of Temperature and Heat in SMA which consists of 50 objective questions with 5 alternative answers (A, B, C, D and E) which were tested on 168 students of class XI from 2 SMA in the city of Perbaungan. The instrument was obtained using the documentation method, namely to obtain a set of questions along with answer keys, student response responses, syllabus and question-making grids.

3.1.1 Analysis Stage

The steps taken at this stage are to observe the research subjects and analyze the material in accordance with this research. The results of the needs analysis based on the results of interviews with physics teachers at SMA Negeri 2 Perbaungan and SMA Setia Budi Abadi Perbaungan that teachers have never conducted an assessment/assessment of student learning outcomes with similar test instruments or high-level knowledge abilities.

3.1.2 Design Stage

This stage is the initial design of the draft of the objective test instrument for High-Level Knowledge Ability material on temperature and heat in SMA. The draft of the instrument is in the form of a grid of test questions, 50 test items, and scoring guidelines. The results of the test questions grid based on material analysis and literature study obtained the results of the instrument grid for high level knowledge of temperature and heat material in high school with characteristics and indicators referring to the dimensions of cognitive processes which are then adjusted to the temperature and heat material and analysis of taxonomic cognitive aspects blooms.

3.1.3 Development Stage

3.1.3.1 Content Validation

Content validation is carried out by reviewing each item of questions prepared with supporting materials such as: (1) Question grids, (2) Test Questions/Instruments, (3) Validation Sheets. The study was carried out to determine the content validation by asking for expert judgment who has competence in their field to assess the accuracy of the content of the items by involving 3 experts. The results of content validation by experts are shown in Table 1.

Table 1. Content Validation Study Results

No	Aspects Understood	Number of Questions That do not Meet The Criteria	Amount
1	Material	1, 2, 3, 4, 5, 9, 11, 17, 24, 29, 30, 39, 49, 50	14
2	Construction	10, 18	2
3	Language/Culture	-	-

3.1.3.2 Construct Validation

Items that have been reviewed by experts are processed using the CVR index with a scale of 1, 2 and 3, where if TR (can be used without revision) is scaled 3, RK (can be used with minor revisions) is scaled 2 and

PK (cannot be used yet, still need consultation) scale. The categories of qualitative test item analysis results are presented in table 2.

Table 2. Construct Validation Study Results

No	Aspects Understood	Amount
1	Accepted	34
2	Revised	14
3	Rejected	2

3.1.3.3. Revision of Expert Review Results

The revision of the questions was carried out based on corrections from experts covering aspects of material, construction, and language. The revision of the qualitative analysis can be seen in Appendix 5. As an example in item number 5 according to experts from the aspect according to experts from the language aspect

where the formulation of questions and options are not clear and do not match the measurement objectives.

3.1.4 Implementation Stage

3.1.4.1. Small Group Trial

This small group trial was conducted after obtaining approval from the experts. This trial was carried out in

class XI MIA 1 SMA Negeri 2 Perbaungan with 35 students. Analysis of questions using the Ms. Excel.

3.1.4.2. Small Group Trial Study Results

3.1.4.2.1 Validity

The validity of the items was calculated using the biserial point formula and obtained 45 valid items and 5 invalid items.

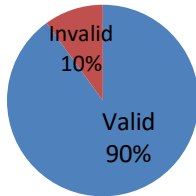


Figure 1. The results of the validity of the small group trial

3.1.4.2.2 Reliability

The reliability of the test in this study was calculated by the Kuder-Richardson formula 20 (KR-20) and it was known that the reliability of the questions developed was 0.891 (high reliability).

3.1.4.2.3 Level of Difficulty

According to Arikunto (2017), "Good questions are questions that are not too easy and not too difficult". Therefore, it can be stated that a good question is a question with a moderate category at $0.3 < p < 0.7$. The results of the level of difficulty can be seen in the following figure.

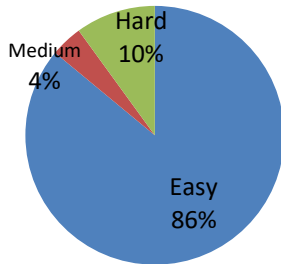


Figure 2. The results of the difficulty level of the trial small group

3.1.4.2.4 Discriminating Power

The analysis of the discriminatory power of this item aims to distinguish between students who have high abilities and students who have low abilities, or to distinguish the upper and lower groups. The image of the discriminatory power results can be seen in Figure 3.

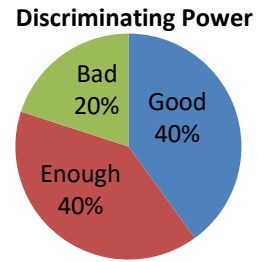


Figure 3. The results of the discriminating power of small group trials

3.1.4.2.5 Effectiveness of the distractor

Sudijono (2017) states that the effectiveness of the distractor can carry out its function properly if the distractor has been selected at least 5% of all test participants. Meanwhile, in terms of turnover, a test is said to be good if the turnover is not more than 10% of students.

In the small group test, there are 35 students, so the distractor will be said to be effective if a minimum of 2 students is selected and the answer is a maximum of 4. The distractor is said to be functioning if all the distractors on each question are functioning well, if the distractor is not functioning or misleading then the distractor needs to be revised. .

3.1.4.2.6 Interpretation of Small Group Trial Results

Based on the data obtained, 41 items (80%) can be accepted and stored in the question bank for the objective test of high-level knowledge on temperature and heat material in high school and tested in large group trials because they have met the validity, difficulty level, discriminating power, and good distractor effectiveness. There are 4 items (8%) that still need to be revised. While the 5 items (10%) that were rejected could not be used because they did not meet any of the criteria for validity, level of difficulty, discriminatory power, and good distractor effectiveness. The results of the interpretation can be seen in Figure 4.

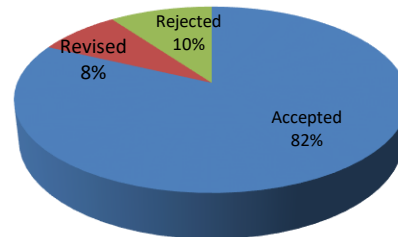


Figure 4. Results of Empirical Analysis of Small Group Trials

3.1.4.3 Revision of Small Group Trial Results

After the small group trial was carried out, a quantitative analysis of the results was carried out regarding the validity, reliability, level of difficulty, discriminating power, and effectiveness of distractors. The results of the analysis obtained that not all instruments meet the criteria for a good test so that there is a need for revisions for questions that are not yet good.

3.1.5 Evaluation Stage

3.1.5.1 Large Group Trial

This large group trial was carried out after obtaining approval from experts. This trial was carried out in class XI MIA of SMA Negeri 2 Perbaungan and XI MIA of SMA Setia Budi Abadi Pebangunan with a total of 133 students.

3.1.5.2 Large Group Trial Study Results

3.1.5.2.1 Validity

Item validity is calculated using the biserial point formula. Based on the analysis of the 45 items, there are 37 valid questions and 8 invalid questions as shown in Figure 5.

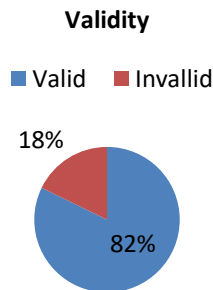


Figure 5. Results of the Validity of the Large Group Trial

3.1.5.2.2 Reliability

The reliability of the test in this study was calculated by the Kuder-Richardson formula 20 (KR-20) and it was known that the reliability of the questions developed was 0.706 (high reliability).

3.1.5.2.3 Level of Difficulty

Items that are classified as good are at a moderate level at $0,3 < p \leq 0,7$. The results of the analysis of the difficulty level of the test instrument from the results of the large group trial obtained 4 items (13,3%) in the easy category ($p > 0,7$), 35 items (75,6%) in the medium category ($0,3 < p \leq 0,7$) dan 6 items (11,1%) in the

difficult category. The results of the level of difficulty can be seen in the following figure 6.

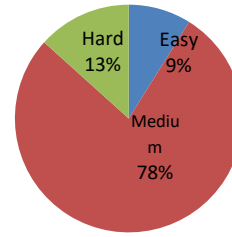


Figure 6. The results of the analysis of the difficulty level of the large group trial

3.1.5.2.4 Discriminatory Power

In large groups, only the two poles are usually taken, namely the top 27% score as the upper group and the lowest 27% score as the lower group. [5]. The image of the results of the discriminatory power can be seen in Figure 7.

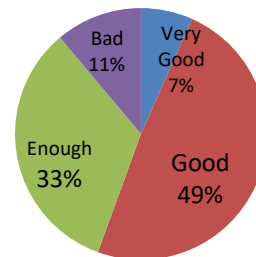


Figure 7. The results of the test discriminatory power analysis big group

3.1.5.2.5 Effectiveness of distractor

The test participants in this large group trial were 133 students. So if calculated 5% of 133 students is 6.65, which means that in this test a distractor is said to be effective if it is chosen by at least 7 students. A good distractor is if it is chosen by many students in the lower group. The distractor is said to be functioning if all the distractors in each item function well, if the distractor is not functioning or is misleading then the distractor needs to be revised.

3.1.5.2.6 Interpretation of Large Group Trial Results

Based on the data obtained, the results of the large group trial received as many as 32 items (71%) and will be stored in the question bank, the questions that need to be revised are 9 items (20%) and the rejected questions are 4 items (9 %) and cannot be used because it does not meet any of the criteria of validity, level of difficulty, discriminatory power and effectiveness of a good distractor. After the 9 questions have been revised, then 41 questions will be stored in the Question Bank for the objective test of high-level knowledge on temperature

and heat material in SMA (Appendix 24). The interpretation results can be seen in Figure 8.

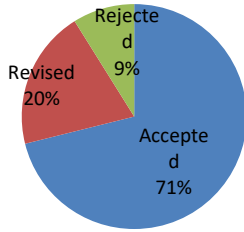


Figure 8. Results of Empirical Analysis of Large Group Trials

3.1.5.3 Revision of Large Group Trial Results

After the large group trial was carried out, a quantitative analysis of the results was carried out regarding the validity, reliability, level of difficulty, discriminating power, and effectiveness of distractors. The results of the analysis obtained that not all instruments meet the criteria for a good test so that there is a need for revisions for questions that are not good.

Evaluation is carried out at the end of each stage of development starting from the assessment design, product manufacture, validation, revision, so that at the evaluation stage the final product is produced. The final product produced is in the form of an Objective Test Question Bank for High Level Knowledge of Temperature and Heat material in SMA.

3.2. Discussion

3.2.1 Problem Analysis Based on Expert Review

The results of the item analysis of the Objective Test of High Level Knowledge Ability on Temperature and Heat Materials in SMA have a very good category with a percentage of 93%, which means that almost all items meet the appropriate criteria. The same thing was also produced by Salamah et al. (2017) where the percentage value of content validity on the feasibility of material, construction and language is 87.5% with a very good category and also with the research of Dian et al. (2018) where the results of the content validation of the development of multiple-choice high order thinking test instruments are in the very valid category with a percentage of 84%.

3.2.2 Question Analysis Based on Public Test

The results of the objective test of the ability of high-level knowledge of temperature and heat material in high school are good. Both in terms of validity, reliability, level of difficulty, discriminatory power and effectiveness of distractors.

Valid items can be reused on the next test or stored in the question bank. Invalid question items are

corrected by adjusting the achievement indicators or adjusting to the technique of preparing the items [10].

A test is said to have high confidence if the test results give a constant result. A good test has a reliability coefficient that ranges from 0.65 to above 0.90[11].

The test discriminatory power in this study averaged 0.331, it means that the distinguishing power of the objective test of the ability of high-level knowledge of temperature and heat material in high school is in the good category [12]. It means that the distinguishing power of the objective test of the ability of high-level knowledge of temperature and heat material in high school is in the good category.

Items that are not good or very bad indicate that the distractor has not been able to function properly, the distractor has no appeal to students who lack mastery of the concept or material. [10].

Based on these data, it is known that 32 items can be accepted and questions that have met all the good test criteria are stored in the question bank for the objective test of high-level knowledge on temperature and heat material in high school because they have met the validity, level of difficulty, discriminatory power, and good distractor effectiveness (Hutapea, et al., 2020). There are 3 questions, namely 3,4, 22, 34, and 38 that do not have effective distractors so they must be revised before being stored in the question bank. Furthermore, the revised questions can be stored in the question bank for the objective test of high-level knowledge on temperature and heat material in high school where the total number of questions has been reduced from 50 to 41 questions.

This study uses classical test theory where the conditions and characteristics of the test takers will affect the results of the study. The conditions in question are internal and external factors of the test takers. Internal factors include the level of intelligence, motivation, health and so on, while the conduciveness of the exam room is one of the external factors that also affects the condition of the test takers. The more conducive the exam room, the better the results will be. The lower the ability of the group of test takers, the more difficult the group of items and vice versa [13]. The discriminatory power depends on the homogeneity of the test takers. This limitation causes differences in results if this study is used in groups of test takers with different characteristics of the object of trial [14].

The results of the study stated that the objective test of the ability of students' high-level knowledge on the material of temperature and heat in high school was in the good category. Repair of questions is needed to improve validity, discriminatory power, level of difficulty and effectiveness of distractors so that all items can be of good value so that the items are of high quality. Good quality items will be able to carry out

their function as a tool for good evaluation. The application of a fine system also needs to be done to minimize the possibility of students guessing the answers to multiple choice questions. Giving a fine with a reduction in value for each wrong answer, students will be more careful in choosing answers. If students really don't know, then students prefer not to answer than just guessing the answer for fear that their score will decrease [5].

5. CONCLUSIONS

The conclusions obtained in the research on developing an objective test of high-level knowledge of temperature and heat material in high school are based on expert reviews in terms of material, construct, and language aspects, the test of high level knowledge of temperature and heat material in high school is in the very good category (92.4 %). The results of the analysis of 45 items in the public test, namely small group trials and large group trials, obtained 32 items that can be accepted and stored in the question bank for high-level knowledge of temperature and heat material in high school that have met the requirements of good test qualifications (standard test). 6 items were rejected and could not be used because they did not meet the criteria of validity, level of difficulty, discriminatory power, and effectiveness of distractors.

ACKNOWLEDGMENTS

First and foremost, I would like to show my deepest gratitude to my lecturers and professors in my university, who have provided me with valuable guidance in every stage of the writing of this thesis. Further, I would like to thank my husband, children, family and all my friends for their encouragement and support. Without all their enlightening instruction and impressive kindness, I could not have completed my thesis.

REFERENCES

- [1] Wagiran, *Inovasi Pembelajaran Masa Depan*. 2007.
- [2] A. and M. G. N. Muslim, development of Reasoning Test Instruments Based on TIMSS Framework for Measuring Reasoning Ability of Senior High School Student on the Physics Concept, *J. Phys. Conf. Ser.*, vol. 2, 2017, doi: 10.1088/1742-6596/812/1/012108.
- [3] Michael Ina, *Michael Allen's Guide to E-learning*. Canada, 2013.
- [4] J. Li and C. Singh, Developing and validating a conceptual survey to assess introductory physics students' understanding of magnetism, *Eur. J. Phys.*, vol. 38, no. 2, 2017, doi: 10.1088/1361-6404/38/2/025702.
- [5] A. Lichtenberger, C. Wagner, S. I. Hofer, E. Stern, and A. Vaterlaus, Validation and structural analysis of the kinematics concept test, *Phys. Rev. Phys. Educ. Res.*, vol. 13, no. 1, pp. 1–13, 2017, doi: 10.1103/PhysRevPhysEducRes.13.010115.
- [6] E. Istiyono, D. Mardapi, and S. Suparno, Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PysTHOTS) Peserta Didik SMA,” *Jurnal Penelitian dan Evaluasi Pendidik.*, vol. 18, no. 1, pp. 1–12, 2014, doi: 10.21831/pep.v18i1.2120.
- [7] H. Hadijah and S. Anggereni, Pengembangan Instrumen Tes Hasil Belajar Kognitif Mata Pelajaran Fisika Pada Pokok Bahasan Momentum Dan Impuls Sma Kelas XI, *J. Pendidik. Fis.*, vol. 4, no. 1, pp. 30–34, 2016.
- [8] D. R. U. Sari, S. Wahyuni, and R. W. Bachtiar, Pengembangan Instrumen Tes Multiple Choice High Order Thinking Padapembelajaran Fisika Berbasis E-Learning Di Sma, *J. Pembelajaran Fis.*, vol. 7, no. 1, p. 100, 2018, doi: 10.19184/jpf.v7i1.7231.
- [9] D. T. Tiruneh, M. De Cock, A. G. Weldeslassie, J. Elen, and R. Janssen, Measuring Critical Thinking in Physics: Development and Validation of a Critical Thinking Test in Electricity and Magnetism, *Int. J. Sci. Math. Educ.*, vol. 15, no. 4, pp. 663–682, 2017, doi: 10.1007/s10763-016-9723-0.
- [10] R. Rahayu and M. Djazari, Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi, *Jurnal Pendidikan Akuntansi Indonesia.*, vol. 14, no. 1, 2016, doi: 10.21831/jpai.v14i1.11370.
- [11] Ehigie, *Psychological Tests and Testing (PSY503)*. Ibadan, 2012.
- [12] F. Kara and D. Çelikler, Development of Achievement Test : Validity and Reliability Study for Achievement Test on Matter Changing, *J. Educ. Pract.*, vol. 6, no. 24, pp. 21–27, 2015.
- [13] A. A. Adeleke and E. O. Joshua, Development and Validation of Scientific Literacy Achievement Test to Assess Senior Secondary School Students ' Literacy Acquisition in Physics, *J. Educ. Pract.*, vol. 6, no. 7, pp. 28–43, 2015.
- [14] E. Y. Yunita, W., Sarwanto., & Ekawati, Pengembangan Tes Fisika SMA Kelas X Semester Ganjil, *J. Pendidik. Fis.*, vol. 1, 2013.