Developing Two-Tier Essay for Diagnostic Test Instrument to Identify Student Learning Difficulty

Bagas Abiyuu Pratama\textsuperscript{a}, Bambang Subali\textsuperscript{a}, Pratiwi Dwijananti\textsuperscript{a}

\textsuperscript{a}Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Central Java, 57174, Indonesia

\textsuperscript{*}Corresponding author: Blulukan 003/006, Colomadu, Karanganyar, Central Java, 57174, Indonesia. E-mail addresses: bagasabiyuupratama@gmail.com

\textbf{a r t i c l e i n f o}

Article history:
Received: 7 May 2018
Received in revised form: 25 July 2018
Accepted: 25 December 2018
Available online: December 2018

Keywords:
Diagnostic test
Two-tier test
Student’s learning difficulties

\textbf{a b s t r a c t}

The purpose of this study was to develop a two-tier essay diagnostic test instrument (DTTE) to identify students’ learning difficulties. The research adopted by Borg and Gall stages with data collection techniques used are interviews and tests. In the development, there were five valid items DTTE with material feasibility (100%), construction (95.5%), and language (100%). The development results showed the instrument reliability criteria using Cronbach’s Alpha (0.76) for the limited trial and field test (0.81). Based on these results DTTE instruments meet the criteria of good, valid and reliable. Analysis of students’ learning difficulties based on learning purpose, they are understanding the kinetic and potential energy, able to explain the influence of angle on work, able to explain the work sense in the graph, and understanding the effort by various styles, to understand relationship work with kinetic energy, with potential energy, and understand conservation of mechanical energy is 92.6%, 90.5%, 89.4%, 89.4%, and 91.5%.

1. Introduction

Science is a collection of experiences and knowledge arranged systematically, where each of these parts is interdependent (Syafie, 2015). The subject of natural sciences serves to provide knowledge about the natural environment and develop skills, insight, and awareness of technology about its use for everyday life (Perwitasari, 2015).

Physics is one of the branches of science where every understanding of the concept is interconnected. Students’ understanding of a concept is very influential in learning other concepts (Siswaningsih, 2014). It is not just knowledge in the form of facts, concepts, and principles but also a learning process that gives students direct experience in understanding the natural surroundings scientifically. Physics learning has a goal to let students think more with a scientific mindset about everything, especially regarding the natural environment (Syafie, 2015).

Students may join the learning process on certain topics well, work on tests with good results, but still do not change their initial ideas that intersect with the topic even if it is
contrary to the scientific concepts taught (Siswaningsih, 2014) or even learning that tends to remember rather than understand the concept of the material being taught. In general, students' understanding of science concepts and science phenomena is part of the key in various science curricula. To measure the effectiveness of learning in facilitating students' understanding of science concepts, assessment tests must be available for teacher use (Adodo, 2013).

Referring to the results of the National Examination held in the 2014/2015 school year, the percentage of the mastery of physics test in Surakarta has the worst percentage of all abilities tested, which was only 41.60%. Therefore, an evaluation is needed by the teacher to fix errors in the learning process and student learning difficulties. Evaluation is one of the factors that play a role in developing students’ potential, the reason why evaluation plays a role in developing students' potential, because evaluation is an important factor to determine the success or failure of the learning process, and at the same time it can influence the subsequent learning process (Fortuna, Chandra & Gloria, 2013). Evaluation has several functions, one of which is diagnostics that function to find out the difficulties of problems faced by students in the learning process/activities. Through diagnostic tests, it can be designed and sought to overcome and or help the person to overcome his difficulties and or solve the problem (Rositasari, Nanda & Salamah, 2014).

Many evaluation techniques are used in measuring the level of understanding and identifying the location of students’ errors in the material “effort and energy.” Combination or one of the mapping, prediction, observation, description concepts, interviews on facts and events, interviews on concepts, word associations, and diagnostic tests are the most widely used techniques (Bayrak, 2013), and in this study diagnostic tests were used to identify the location of students’ errors. Diagnostic tests are a device of concentration which focuses on student learning difficulties that repeat over and over, where the errors have not been resolved and cause learning difficulties in students (Gurel et al., 2015). Diagnostic tests are a set of instruments that can be used by student misconceptions (Kurniawan & Maryanti, 2018). One diagnostic technique that can be used is a two-tier diagnostic test (TDTT), the two-tier multiple choice diagnostic test consists of two tiers, namely, the first tier contains answers to questions and in the second tier contains the reasons for the answer to the first part (Kanli, 2015; Chandrasegaran et. al., 2007). In the second tier, students must write about the reasons for answering in the first tier, this is intended to see whether the respondent can only work on
the questions because he worked on similar questions and memorized steps and formulas that were used or correct because of random answers or respondents understood well about the concepts presented in the question/test.

In general, TDIT is in the form of multiple choice in the first tier and reasoning for the second tier. As it expressed by Kurniasih and Nukhbatul (2017), the category of student understanding is determined from the right or wrong answers in answering each tier of questions with the answers provided, but in the multiple-choice questions (MCQ), the teachers are unable to see the flow of problem-solving by students so that students' accuracy and small mistakes that cause fatal answer errors cannot be observed. Unlike the essay question, where the answer choices are not provided and students must answer with sentences, so they can train students to convey information verbally. Besides, the essay exam also requires a better understanding of science and can be used to measure the level of one’s understanding of a science deeper (Fitri & Asyikin, 2015). The form of essay questions is very helpful for students to be able to maximize all knowledge they have in writing to answer the questions asked. Compared to other forms (multiple choice, true-false, etc.) this form is very flexible (Siswanto, 2006). For this reason, a two-tier diagnostic essay is developed so students can convey their knowledge maximally. Besides, two-tier essay diagnostic instruments can see the flow of answers from students so that it will reveal whether students solve the problem by remembering the completion of the questions given by the teacher, or students completing the questions from the results of the interpretation of the questions into their understanding.

Through this research on the development of diagnostic two-tier essay test (DTTE) instruments, it is expected to make public references about the steps that can be taken to produce instrument products that can be used to diagnose student learning difficulties more accurately.

2. Methods

The research design applied in this study is Research and Development (R & D) by adopting the stages of Borg and Gall (1996) in the Educational Research book: An Introduction (4th ed) which includes 10 stages namely (1) information gathering (2) planning (3) developing the initial product form (4) initial trial (5) revising of the initial trial (6) field trial (7) refining the product results from field trials (8) operational field trials (9) final product improvement (10) disseminating and implementing the products. But in its
implementation, Borg and Gall (1996) suggested only to step 7 because it is a series of formative evaluations while steps 8 to 10 are a series of summative evaluations. Data analysis techniques used are descriptive qualitative and quantitative. Qualitative data was obtained from the validator by two reviewers namely the teacher, observation at the preliminary study stage and the results of the interview. The data of the review by the validator aims to determine the feasibility of the instrument products to be tested on high school students in class XI and the corrections given to accomplish the instrument products. Meanwhile, observation data serves to obtain information about problems affecting the education sector, and interview data used to the results of the analysis of student learning difficulties tested with instrument products. Quantitative data were obtained from the initial results with 64 students from SMAN 2 Surakarta and field trials with 94 high school students from SMAN 1 Surakarta, SMAN 4 Surakarta, and MAN 1 Surakarta. The data obtained is then formulated in the Cronbach's Alpha equation to determine the reliability value of the items which are then interpreted in the reliability category (Sugiyono, 2013).

Descriptive analysis is applied to analyze quantitative data. There are three sources of quantitative data in this study, namely: expert test, initial trial, and field test. Quantitative data from expert test results were analyzed using equations (1)

\[ P = \frac{S}{N} \times 100\% \]  ...(1)

Notes:
P: percentage
S: score
N: maximum score
The percentage results are then interpreted based on the criteria of classification according to Akbar and Sriwiyan (2010).

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Criteria</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>86% - 100%</td>
<td>Strongly Valid</td>
<td>Applied without revision</td>
</tr>
<tr>
<td>70% - 85%</td>
<td>Valid</td>
<td>Applied with minor revision</td>
</tr>
<tr>
<td>60% - 69%</td>
<td>Less Valid</td>
<td>Recommended not to apply</td>
</tr>
<tr>
<td>00% - 50%</td>
<td>Not Valid</td>
<td>Unable to apply</td>
</tr>
</tbody>
</table>

(Akbar & Sriwiyan, 2010)
Quantitative data from the results of the initial trial and the results of field trials were analyzed using the Cronbach's Alpha equation to determine the reliability of the items. The minimum Cronbach Alpha value that can be accepted and used for research is 0.60 (Hair et al., 2010). Analysis of the data on the results of the initial trial will be used to find out the criteria of the questions used, namely reliability to determine whether DTTA instruments need to be revised and the difficulty level of the questions fulfill the criteria to be achieved. DTTA instruments are targeted to have a moderate level of difficulty. According to Suharsimi (2013), good questions are not too easy because the problems that are too easy do not stimulate students to enhance their problem-solving efforts. Meanwhile, problems that are too difficult will discourage students and not have the enthusiasm to try again because they are beyond the capability. While field test reliability is used to confirm that DTTA instruments are reliable.

Analysis of the data used in processing the initial test data and field tests to get the Cronbach Alpha value is with the help of SPSS software. The Alpha Cronbach DTTA instrument obtained is then interpreted into the instrument reliability criteria according to Fleiss (1981).

<table>
<thead>
<tr>
<th>Alpha Cronbach</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75 – 1.00</td>
<td>Strongly High</td>
</tr>
<tr>
<td>0.60 – 0.74</td>
<td>High</td>
</tr>
<tr>
<td>0.40 – 0.59</td>
<td>Enough</td>
</tr>
<tr>
<td>0.00 – 0.39</td>
<td>Bad</td>
</tr>
</tbody>
</table>

(Widiarso, 2007)

3. Results and Discussion

3.1 The Development of DTTA instruments

The development of diagnostic instruments is in the form of test questions for students. This two-tier essay diagnostic test covers students' interpretation abilities in graphics and images. The form of the DTTE instrument is a matter of reasoning essay with the composition of the questions shown in Table 3.
Table 3 The Composition of diagnostics test for Two-tier Essay Test

<table>
<thead>
<tr>
<th>Types of Interpretation Capability</th>
<th>Small Scale Test</th>
<th>Big Scale Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Question</td>
<td>Percentage (%)</td>
</tr>
<tr>
<td>Graphic Interpretation</td>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>Figure Interpretation</td>
<td>3</td>
<td>60</td>
</tr>
</tbody>
</table>

The research data consisted of expert validation data, initial trial data, and field trial data. Expert validation data was obtained through a closed questionnaire by experts consisting of three aspects, namely: substance aspects, construction aspects, and language aspects. The results obtained by each product aspect by experts are shown in Table 4.

Table 4 Recapitulation of Question Validity by Experts

<table>
<thead>
<tr>
<th>Validator</th>
<th>Materials</th>
<th>Construction</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score</td>
<td>Percentage (%)</td>
<td>Score</td>
</tr>
<tr>
<td>V-1</td>
<td>16</td>
<td>100</td>
<td>11</td>
</tr>
<tr>
<td>V-2</td>
<td>16</td>
<td>100</td>
<td>12</td>
</tr>
<tr>
<td>Average</td>
<td>16</td>
<td>100</td>
<td>11,5</td>
</tr>
<tr>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strongly Valid</td>
<td></td>
<td>Strongly Valid</td>
</tr>
</tbody>
</table>

Based on the data analysis aspects of the validity of the experts stated that the question was feasible in testing on high school students of class XI. Tests on students carried out in two stages, namely: the initial test phase and the field trial stage. The results of the initial trial show the criteria for the question which will be used to determine the next step. Based on the data analysis, the reliability score and the difficulty level of the questions are shown in Table 5.

Table 5. Recapitulation of the Small Scale Test for Level of Difficulty

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Difficulty Index</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.544</td>
<td>Moderate</td>
</tr>
<tr>
<td>2</td>
<td>0.598</td>
<td>Moderate</td>
</tr>
<tr>
<td>3</td>
<td>0.498</td>
<td>Moderate</td>
</tr>
<tr>
<td>4</td>
<td>0.537</td>
<td>Moderate</td>
</tr>
<tr>
<td>5</td>
<td>0.509</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

Based on the data analysis of the level of difficulty test items in the small-scale trial phase, all items on average have a difficulty index of 0.5, or it can be concluded that all items in the problem are at the level of moderate difficulty. The use of real questions with moderate
difficulty level aims to stimulate the enthusiasm of students in working on the problem because the problem is not too difficult.

The results of the question reliability analysis show the value of Cronbach Alpha (0.76), this shows that in the initial trial the instrument was reliable. Based on the results of the analysis at the initial trial phase, there were some drawbacks of the instrument. Namely, the processing time was too fast, and some images were slightly less representative. After revision, DTTA's instruments are ready to be field tested. Based on the analysis, the Cronbach Alpha value is obtained (0.81), indicating that the DTTE instrument is reliable

Based on the research with the DTTE test, the comparison of students with the correct answers in the first tier and second tier is shown in Table 6.

Table 6. Recapitulation of Percentage of Student Answers

<table>
<thead>
<tr>
<th>Question Number</th>
<th>The Percentage of Correct Answer in the First Tier</th>
<th>The Percentage of Correct Answer in the Second Tier</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>43.6 %</td>
<td>11.3%</td>
</tr>
<tr>
<td>2</td>
<td>67.1%</td>
<td>14.7%</td>
</tr>
<tr>
<td>3</td>
<td>69.8%</td>
<td>13.8%</td>
</tr>
<tr>
<td>4</td>
<td>35.1%</td>
<td>22.8%</td>
</tr>
<tr>
<td>5</td>
<td>39.3%</td>
<td>18.3%</td>
</tr>
</tbody>
</table>

Based on Table 6, the percentage of answers in the second tier is lower than the second tier, this shows that students still have difficulty in answering questions that use their comprehension abilities and tend to use their memory skills in answering questions, and when students continue to use memory in answering questions and not understanding what he learned, the students will find it difficult for students to meet questions with the types of questions that they have never encountered (Anderson & Krathwohl, 2010). This is by the results of the study which states (Nofiana, Sajidan & Puguh, 2014) that two-tier questions are more challenging than ordinary questions and the two-tier questions are more able to measure and improve students' thinking skills. The profile of student learning difficulties analyzed based on the learning objectives approach can be seen in Graph 1.
T1 in Graph 1 shows the first learning goal, namely, understanding kinetic and potential energy (gravity and springs), T2 (able to explain the influence of angles on effort concepts, explaining effort understanding in graphs, and understanding effort by various styles), T3 (able to understand relationships between effort and kinetic energy), T4 (able to understand the relationships between effort and potential energy), and T5 (able to understand the conservation of mechanical energy).

Based on Graph 1, many students have difficulty learning in the effort and energy materials. The average student error lies in understanding weak concepts and based on the analysis of student answers in working on DTTE instruments shows there are still many student answers that are wrong because they are inaccurate and some findings of small mistakes made by students namely, in answering questions students tend to answer questions by entering the formula they remember without understanding the intent of the problem. For example, in looking for ‘effort’ values that should be obtained from the area of the graph students immediately look for them by entering the formula $W = Fs$ and from the results seen (Figure 2 with the aim of T1) students only come from the numbers and incidentally the results are the same as the answer key, the results shows the ability to interpret the bad and students are too fixated on the formula.
Also, students have difficulty in interpreting graphs (Figure 3), this is seen when students cannot draw effort graphics, and students cannot determine the range of variables needed to answer questions.

4.

Two objects of equal mass drawn by the force $F = 12N$ in $X = 6m$.

A. Do object 1 and object 2 subject get the same effort? If they do not, which one get bigger effort? ($\tan \alpha = \sqrt{3}$)

B. Draw graphic of the effort in object 1 and 2 in an $F$ graphic to $X$

Interpretation ability is one of the important aspects in understanding science. According to Mustain (2015), the ability to understand graphics is a capability that must be possessed and becomes a very important part because interpretation is part of the complementary core experiments of science. The students' trigonometric abilities are also still weak which can be
seen in the students’ answers to the questions that explain the influence of angles in the ‘effort’ concept.

Based on the results of diagnostic tests there are still many students who have difficulty in learning, especially in the ability to interpret the understanding that they received with the problems they faced. This is seen when students only match the formula in solving the problem. Hence, the teacher needs to evaluate the way of learning so that learning in understanding is comprehensive. It is true that what the students do is not only remembering, but also how to solve the problem so that students will not be confused when facing new problems. Students’ difficulties in solving essay problems are caused by a lack of practice in the form of analysis (Pratiwi & Lepiyanto, 2017; Lucariello, Tine & Ganley, 2014). Kowalski and Taylor (2004) provide a solution so that students do not experience difficulties, namely familiarizing students with critical thinking in learning.

4. Conclusion

The results of this study can be concluded that DTTE instruments can be developed with the stages of Borg & Gall to detect student learning difficulties and to obtain reliable instrument products with Cronbach Alpha characteristics in limited trials (0.76) and field trials (0.81). Profile of student learning difficulties based on learning objectives, understanding kinetic and potential energy (gravity and springs), being able to explain the influence of angles on ‘effort’ concepts, explaining ‘effort’ understanding in graphics, and understanding ‘effort’ by various styles, being able to understand the relationships between effort and kinetic energy, being able to understand the relations between effort and potential energy, and able to understand the eternity of mechanical energy respectively 92.6%, 90.5%, 89.4%, 89.4%, and 91.5%.

Acknowledgments

We thank you to Drs. Bambang Budi Hartono, as a physics teacher at SMAN 1 Surakarta; Agus Nugroho, S.Pd., as the physics teacher of MAN 1 Surakarta, Drs. Agus Hartanto, M.Hum., as the teacher of Surakarta Public High School 2; Edi Susilo, S.Pd., as the physics teacher of SMAN 1 Magelang; along with all parties who took part in this research.
References


