



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

Bagas Abiyyu Pratama^{a*}, Bambang Subali^a, Pratiwi Dwijananti^a

^a Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Central Java, 57174, Indonesia

*Corresponding author: Blulukun 003/006, Colomadu, Karanganyar, Central Java, 57174, Indonesia. E-mail addresses: bagasabiyyupratama@gmail.com

article info

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

abstract

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%.

2018 Scientiae Educatia: Jurnal Pendidikan Sains

1. Introduction

Science is a collection of experiences and knowledge arranged systematically, where each of these parts is interdependent (Syafiie, 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 (Syafiie, 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, TDDT 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)

$$P = \frac{S}{N} \times 100\% \quad \dots(1)$$

Notes:

P: percentage

S: score

N: maximum score

The percentage results are then interpreted based on the criteria of classification according to Akbar and Sriwiyana (2010).

Table 1. Eligibility criteria for two-tier diagnostic instruments for test subjects of effort and energy materials for high school students of class XI

Percentage	Criteria	Mean
86% - 100%	Strongly Valid	Applied without revision
70% - 85%	Valid	Applied with minor revision
60% - 69%	Less Valid	Recommended not to apply
00% - 50%	Not Valid	Unable to apply

(Akbar & Sriwiyana, 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 2. Reliability criteria according to Fleiss

Alpha Cronbach	Criteria
0,75 – 1,00	Strongly High
0,60 – 0,74	High
0,40 – 0,59	Enough
0,00 – 0,39	Bad

(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

Types of Interpretation Capability	Small Scale Test		Big Scale Test	
	Number of Question	Percentage (%)	Number of Question	Percentage (%)
Graphic Interpretation	2	40	2	40
Figure Interpretation	3	60	3	60

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

Validator	Materials		Construction		Language	
	Score	Percentage (%)	Score	Percentage (%)	Score	Percentage (%)
V-1	16	100	11	91	20	100
V-2	16	100	12	100	20	100
Average	16	100	11,5	95	20	100
Notes	Strongly Valid		Strongly Valid		Strongly Valid	

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

Question Number	Difficulty Index	Category
1	0,544	Moderate
2	0,598	Moderate
3	0,498	Moderate
4	0,537	Moderate
5	0,509	Moderate

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.

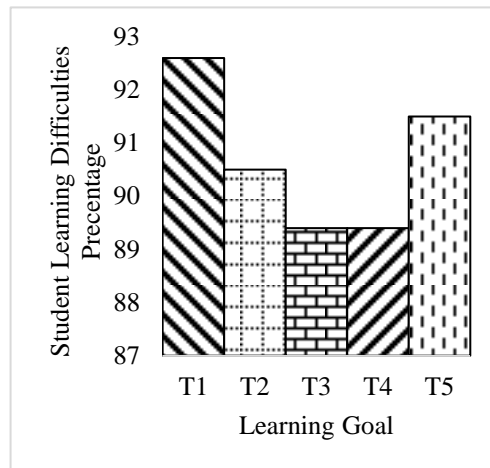
The results of the question reliability analysis show the value of *Cronbach Alpha* (0.76), this shows that in the initial trial the instrument was reliable. Based on the results of the analysis at the initial trial phase, there were some drawbacks of the instrument. Namely, the processing time was too 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

Question Number	The Percentage of Correct Answer in the First Tier	The Percentage of Correct Answer in the Second Tier
1	43.6 %	11.3%
2	67.1%	14.7%
3	69.8%	13.8%
4	35.1%	22.8%
5	39.3%	18.3%

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.



Graphic 1. Recapitulation of Student Learning Difficulties Profile

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).

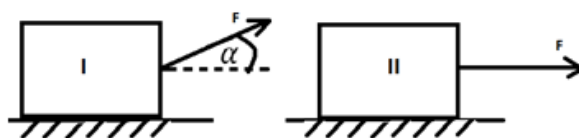
Based on Graph 1, many students have difficulty learning in the effort and energy materials. The average student error lies in understanding weak concepts and based on the analysis of student answers in working on DTTE instruments shows there are still many student answers that are wrong because they 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.

$W = F \cdot S$ $= \frac{(5+1) \cdot 10}{2}$ $= 30$	$W = \Delta E_k$ $30 = \frac{1}{2} m (V_2^2 - V_1^2)$ $30 = \frac{1}{2} \cdot 12^6 (V_2^2 - 4)$ $30 = 6 (V_2^2 - 4)$ $5 = V_2^2 - 4$ $9 = V_2^2$ $V_2 = 3 \text{ m/s}$
(a)	(b)

Figure 2 Questions and Answers of DTTE Instruments with the aim of T1

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 $F = 12\text{N}$ in $X = 6\text{m}$,

- 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

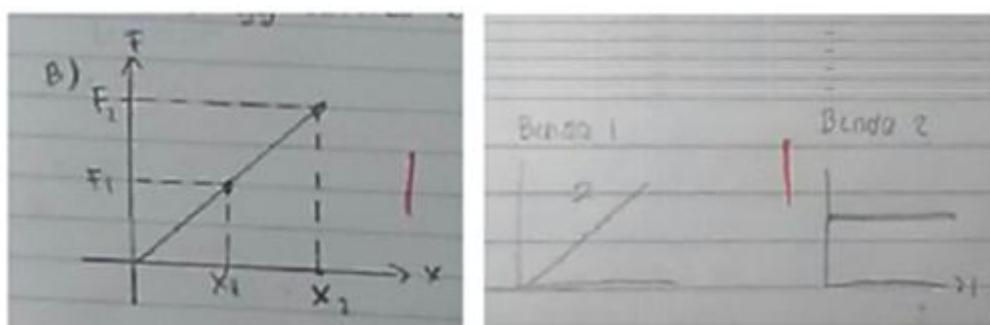


Figure 3 Questions and Answers of DTTE Instruments with the aim of T2

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

- Adodo, S. O. (2013). Effects of two-tier multiple choice diagnostic assessment items on students' learning outcome in basic science technology (BST). *Academic Journal of Interdisciplinary Studies*, 2(2), 201-210. DOI: 10.5901/ajis.2013.v2n2p201
- Akbar, S., & Sriwiyana, H. (2010). Pengembangan kurikulum dan pembelajaran Ilmu Pengetahuan Sosial (IPS). Yogyakarta: Cipta Media.
- Anderson, L. W., & Krathwohl, D. R. (2010). Kerangka landasan untuk pembelajaran, pengajaran, dan asesmen. Yogyakarta: Pustaka Pelajar.
- Bayrak, B. K. (2013). Using two-tier test to identify primary students' conceptual understanding and alternative conceptions in acid-base. *MIJE*, 3(2), 19-26. DOI: 10.13054/mije.13.21.3.2
- Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, 8(3), 293-307. DOI: 10.1039/B7RP90006F
- Fitri, R., & Asyikin, N. A. (2015). Aplikasi penilaian ujian essay otomatis menggunakan metode cosine similarity. *Jurnal Poros Teknik*, 7(2), 88-94. DOI: 10.31961/porosteknik.v7i2.218
- Fortuna, D., Chandra, E., & Gloria, R. Y. (2013). Pengembangan tes diagnostik untuk mengukur miskonsepsi siswa pada pokok bahasan sistem regulasi manusia untuk siswa SMA kelas XI semester II. *Scientiae Educatia: Jurnal Pendidikan Sains*, 2(2). Available online: <http://www.syekhnurjati.ac.id/jurnal/index.php/sceducatia/article/view/514>
- Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction*. Longman Publishing.
- Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(5), 989-1008. DOI: 10.12973.eurasia.2015.1369a
- Hair, J. F. J., Black, W. C., Babin, B. J., & Andersen, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River: Prentice Hall
- Kanli, U. (2015). Using a two-tier test to analyses students' and teachers' alternative concepts in astronomy. *Science Education International*, 26(2), 148-165. Available online: <https://files.eric.ed.gov/fulltext/EJ1064041.pdf>
- Kowalski, P., & Taylor, A. K. (2004). Ability and critical thinking as predictors of change in students' psychological misconceptions. *Journal of Instructional Psychology*, 31(4). Available online: <https://www.questia.com/library/journal/1G1-126582634/ability-and-critical-thinking-as-predictors-of-change>
- Kurniasih, N., & Nukhbatul, B. H. (2017). Penggunaan tes diagnostik two-tier multiple choice untuk menganalisis miskonsepsi siswa kelas X pada materi archaeobacteria dan eubacteria. *Jurnal Tadris Pendidikan Biologi*, 8(1), 114-127. Available online: <http://ejournal.radenintan.ac.id/indek.php/biosfer/index>
- Kurniawan, D., & Maryanti, S. (2018). Analysis of math teacher candidates' misconception on the dynamic electricity concept. *Scientiae Educatia: Jurnal Pendidikan Sains*, 7(1), 67-78. DOI: 10.24235/sc.educatia.v7i1.2481
- Lucariello, J., Tine, M. T., & Ganley, C. M. (2014). A formative assessment of students' algebraic variable misconceptions. *The Journal of Mathematical Behavior*, 33, 30-41. DOI: 10.1016/j.jmathb.2013.09.001

- Mustain, I. (2015). Kemampuan membaca dan interpretasi grafik dan data: studii kasus pada siswa kelas 8 SMPN. *Scientiae Educatia: Jurnal Pendidikan Sains*, 5(2). Available online: <http://www.syekhnurjati.ac.id/jurnal/index.php/sceducatia/article/view/493>
- Nofiana, M., Sajidan, & Pugh. (2014). Pengembangan instrumen evaluasi two-tier multiple choice question untuk mengukur keterampilan berpikir tingkat tinggi pada materi kingdom plantae. *Jurnal Inkuiri*, 3(2), 60-67. Available online: <http://jurnal.fkip.uns.ac.id/index.php/sains>
- Perwitasari, A. D. (2015). Pengembangan tes diagnostik berbasis web pada materi termodinamika untuk mengidentifikasi tingkat pemahaman konsep siswa. *Skripsi*. Semarang: Universitas Negeri Semarang. Available online: <http://lib.unnes.ac.id/21894/>
- Pratiwi, D., & Lepiyanto, A. (2017). Profil miskonsepsi calon guru biologi pada mata kuliah PPL II. *BIOEDUKASI (Jurnal Pendidikan Biologi)*, 8(1), 33-44. DOI: 10.24127/bioedukasi.v8i1.834
- Rositasari, D., Nanda, S. D., & Salamah A. (2014). Pengembangan tes diagnostik two-tier untuk mendeteksi miskonsepsi siswa SMA pada topik asam-basa. *Edusains*, 6(2), 170-176. Available online: <http://journal.uinjkt.ac.id/index.php/edusains/article/view/1148>
- Siswaningsih, W. (2014). Pengembangan tes diagnostik two-tier untuk mengidentifikasi miskonsepsi pada materi kimia siswa SMA. *Jurnal Pengajaran MIPA*, 19(1), 117-127. Available online: http://journal.fpmipa.upi.edu/index.php/jpmipa/article/download/487/pdf_21
- Siswanto. (2006). Penggunaan tes essay dalam evaluasi pembelajaran. *Jurnal Akuntansi Pendidikan Indonesia*, 5(1), 55-61. Available online: <https://journal.uny.ac.id/index.php/jpakun/article/view/864>
- Sugiyono. (2013). *Metode penelitian pendidikan pendekatan kualitatif, kuantitatif, dan R & D*. Bandung: Alfabeta.
- Suharsimi, A. (2013). *Dasar-dasar evaluasi pendidikan*. Jakarta: PT Bumi Rupa Aksara.
- Syafiie, I., K. (2015). *Pengantar ilmu pemerintahan*. Bandung: PT Refika Aditama.
- Widiarso, W. (2007). Mengestimasi reliabilitas. Diperoleh 3 Juni 2013, Available online: http://widhiarso.staff.ugm.ac.id/files/bab_2_estimasi_reliabilitas_via_spss.pdf