Analysis of student's learning outcomes test questions Department of Automotive Engineering FT UNP

Hendra Dani Saputra[1]*, Sartika Anori[2], Donny Fernandez[3], Dedi Setiawan[4]

[1,3,4,] Jurusan Teknik Otomotif, Universitas Negeri Padang
Jl. Prof. Dr. Hamka Kampus UNP Air Tawar Padang, Indonesia-25132

[2] Jurusan Teknik Elektronika, Universitas Negeri Padang
Jl. Prof. Dr. Hamka Kampus UNP Air Tawar Padang, Indonesia-25132

*Corresponding author: hendradani@ft.unp.ac.id

Abstract

This study is a follow-up to the results of previous studies, which stated that e-learning media did not have a significant effect on improving learning outcomes. One way of measuring learning outcomes is by presenting test questions on E-Learning media. So that a good test item will be able to provide a good measurement of learning outcomes. This research is included in quantitative descriptive research. The research aims to see the level of validity, reliability, difficulty index and Different power test questions for students of the Department of Automotive Engineering, FT UNP. The research method presenting test items in the form of multiple choice. Then proceed with quantitative analysis of the test items. The results of the analysis will provide an overview of the criteria for valid items of 84% and the reliability index of the questions in the criteria of 0.899. The results of the test difficulty index analysis, that 80% of the test questions are included in the category of good test items. As many as 20% of the test questions including the category are too easy. The results of the analysis of the different power of the test items are 28% including the bad classification. Classification test questions are sufficient at 40%. Classification of different test items is good at 32%. The category of good test items is a category of test items with sufficient and good classification.

Keywords: Test items, analysis, validity, reliability, difficulty index, different power

1. Introduction

The learning process has the goal of producing good, effective and efficient learning outcomes. The learning process requires appropriate conditions and situations so that good learning outcomes will be achieved.[1] One form of achievement of learning outcomes in the form of the effectiveness of the learning process so as to produce good learning outcomes. The Big Indonesian Dictionary defines effectiveness as usability, activity and the existence of conformity in an activity between someone who carries out a task with the goal to be achieved. Effectiveness can be done with various learning models as an effort to improve student learning outcomes.[2] The effectiveness of learning outcomes includes the realm of knowledge, skills and attitudes that must be possessed by students. Effective learning is influenced by internal factors and external factors.[3] The current condition of learning that is carried out online, in principle, still demands good student learning outcomes. The end of online learning also requires good learning outcomes. The situation and conditions of learning both face-to-face and online must be the same in producing the learning outcomes to be achieved.[4] One of the efforts to get good learning outcomes during the COVID-19 pandemic is by giving test questions online in the form of multiple choice questions. The advantage of multiple choice questions is that they are easy to correct and analyze.[5]. The end of the semester exam is one of the methods used to determine the extent of students' understanding in learning, if the results are maximum, the learning activity will be more effective and vice versa One form of implementing the end of semester exam is by giving test questions. Test questions are one form of measurement in education or learning. Measurement is a procedure for applying numbers or symbols to the attributes of an object or activity in accordance with certain rules.[6]

Learning outcomes measurement activities are carried out in the form of providing numbers based on predetermined criteria. Measurement of learning outcomes is a process of measuring the achievement of competence by students. There are two factors that influence the measurement of learning outcomes, namely internal and external factors. Provide an answer that internal factors are one of the factors that affect learning outcomes. In addition, external factors cannot be ruled out in terms of measuring learning outcomes. Previous research that has been published has concluded that the given E-Learning media does not have a significant effect on learning outcomes. [7] Learning outcomes obtained from giving test questions. Giving test questions is one of the efforts to determine the competence of students. The test is used by an educator in an effort to find out the extent to which the ability of students to master the learning material.

Various forms of measurement/assessment of learning outcomes in the world of education still include aspects of cognitive, psychomotor and affective. The types of measurement/assessment of learning outcomes include written tests, oral tests, and assignments or projects within a certain period of time. Tests usually carried out by educators aim to assess student achievement in the form of daily tests, midterm exams , and final semester exams.[8] Item analysis is an important step that must be taken by educators after the teacher provides an evaluation of the learning outcomes test. The criteria for the questions to be said to be good, the questions that must be possessed include: High validity if the test used can measure what is intended to be measured accurately and Reliability if the test is tested many times which is done at different times and shows a reliable determination.[9]

Analysis of test quality is a step that must be taken to determine the degree of test quality, both as a whole and the items that are part of the test. The test analysis aims to (1) find out the strengths and weaknesses of the questions so that selection and revision actions can be carried out, (2) provide complete information on item specifications, (3) find out the problems contained in the items, (4) to assess the stored items. in a collection of questions or question banks, and (5) is used as information for compiling items.[10] A good test question is to have high validity, reliability or a consistent degree of a test that provides information about the right measurement and is in accordance with what is intended to be measured.

Achievement of learning outcomes can be given through exams or tests. The learning outcomes test given needs to be analyzed for the validity and reliability of the questions. The results of the analysis are needed to produce good test questions. So as to measure the ability and knowledge of students well. The test is one of the efforts to determine the competence of students.The test is used by educators in an effort to measure the extent to which students' abilities in mastering learning materials. There are three forms of tests that can be done by educators.

1. Written test, a test that requires written answers.
2 Oral test, demanding answers verbally.
3. Action test, demands an answer in the form of action

The written test consists of 1) multiple choice questions, 2) filling in, 3) short answers (short), 4) true-false (B-S), 5) matching, and 6) description. During the Covid-19 pandemic, the questions that were often used during the final semester exam were multiple choice questions. Multiple choice test questions do not provide an opportunity to provide a graded assessment because the answers to the questions are clear and definite. This question is used to determine student learning outcomes in terms of knowledge that is memory, understanding, application, analysis, synthesis, and evaluation.

The advantages and disadvantages of multiple choice tests are as follows.

Advantages of Multiple Choice Questions

1. Can measure both recall and high mental processes
2. Can be scored quickly, both by teachers, machines, operators and efficiently.
3. The level of difficulty can be measured.
4. Appropriate for diagnostic tests.
5. Easy to answer students.

Disadvantages of Multiple Choice Questions

1. Difficult to make.
2. Tend to make recall questions.
3. Multiple choice objective test is time consuming.
4. Often ambiguous writing occurs.[5]

Focus on giving test questions with multiple choice types. Multiple choice test is one of the choices that is assumed to be good for measuring learning outcomes. Therefore, an analysis of the level of validity and reliability of the questions was carried out. So that the resulting test questions will be good. A good level of test items has a high validity and a high level of reliability or a high degree of consistency in a test.

2. Method

This research is quantitative descriptive. Research that provides an overview in the form of numbers or statements that are analyzed and assessed statistically. The method used in this research is in the form of giving a test in the form of multiple choice questions. Multiple choice questions that have five answer choices. The results of the answers to the test questions given by students will be tested for data analysis. Data analysis was carried out in the form of validity analysis and reliability analysis of test questions. The results of the analysis will provide an overview of the level of validity and reliability of the test questions. The following is an overview of the research carried out:

Test Questions → Data analysis technique → Validity test / Reliability Test → Valid and Reliable Test Questions
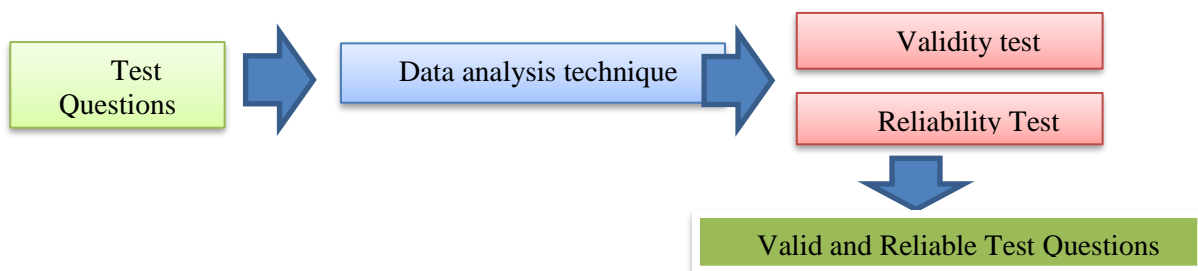
The instrument or data collection tool is in the form of objective test questions to measure cognitive learning outcomes. Objective test questions are given in the form of five answer choices. The test questions tested are adjusted to the material provided during the teaching and learning process. Data analysis techniques include: Validity test and reliability test.

Validity test
Measuring the validity or accuracy of multiple choice questions is formulated using the Biserial I Point correlation, as follows:

$$rpbis = \frac{Mp - Mt}{SDt} \sqrt{\frac{p}{q}}$$

r_{pbis} = coefisien biserial
M_p = the average value of the subjects who answered correctly
M_t = average total score
SD_t = standard deviation
P = the number of students answered correctly $\left(P \frac{\text{many students right}}{\text{total number of students}} \right)$
q = number of students answered incorrectly (q=1 − p)

Validity value of the question by looking at the Serial Point value. Analysis data compared to the value of r_{tabel} product moment significant value 5% . When $r_{pbis} \geq r_{tabel}$ so the item test will be valid.[9]

Reliability Test
Measuring the extent to which a test question for learning outcomes can be trusted, consistent and stable anytime, anywhere and by whom the test is carried out, examined, and assessed. To determine the reliability of the test results, the Single Test-Single Trial approach was used using the Kuder Richardson formula with the code KR21

23

$$r11 = \left(\frac{n}{n-1}\right)\left(1 - \frac{Mt(n-Mt)}{(n)(St^2)}\right)$$

$$Mt = \frac{\sum Xt}{N}$$

$$St2 = \frac{\sum Xt^2 - \left[\frac{\sum Xt}{N}\right]^2}{N}$$

r^{11} = Test reliability coefficient
n   = Number of items
M   = Average test score
St2   = Total variance
N   = Number of participants
Criteria
0,00 – 0,20   = very low reliability
0,21 – 0,40   = low reliability
0,41 – 0,70   = sufficient reliability
0,71 – 0,90   = high reliability
0,91 – 1,00   = very high reliability[9]

Difficulty Index

The difficulty index is a number that shows the difficulty and ease of a problem. To determine the index of difficulty can be determined by the formula:

$$P = \frac{B}{JS}$$

P = Item difficulty index number
B = The number of students who answered the question correctly on the item
JS = Number of students who took the learning outcomes test. [11]

Difficulty index classification of questions:
P = 0.00 – 0.30 = too difficult
P = 0.31 – 0.70 = moderate
P = 0.71 – 1.00 = too easy

Different Power

Different power can be used to look at the ability of a matter that can distinguish between students who have high or low ability. To determine the different power of the test used the formula is:

$$D = PA - PB \qquad \text{with } PA = \frac{B_A}{J_A} \text{ dan } PB = \frac{B_B}{J_B}$$

Information : :
D = Item difference power index number
PA = Proportion of upper group testees who can correctly answer the item in question
PB = Proportion of lower group testees who can correctly answer the item in question
BA = The number of upper group testees who can correctly answer the item in question.
BB = The number of lower group testees who can correctly answer the item in question.
JA = Number of testees included in the upper group
JB = Total number of testees included in the upper and lower groups kelompok

Classification of different power:

D = 0.00 – 0.20 = bad
D = 0,21 – 0,40 = enough
D = 0,41 – 0,70 = good
D = 0,71 – 1,00 = very good. [12]

The population in this study were students D3 majoring in automotive engineering Faculty of Engineering, Universitas Negeri Padang on Vehicle Body Construction courses.

3. Results and Discussion

Validity Analysis Results
The results of the analysis of the validity of the test items are described in the following table:

Table 1.   Validity Analysis Results

Number of Questions	Valid Test Questions	Invalid Test Questions	Valid Percentage	Invalid Percentage
25 Question	21 Question	4 Question	84%	16%

Based on the table data from the results of the validity analysis, it can be seen that the level of validity of the questions is very high. The test questions given are 25 questions. There are 21 valid questions with a percentage level of 84% of the test questions given are valid. Invalid questions are 4 items, with a percentage of 16% invalid questions. The level of valid test questions indicates that the provision of test questions is good. Invalid test questions need to be studied again about the test questions. Invalid test questions are corrected and grammatical revisions are made, answers to test questions or new questions are made that represent the test questions so that the valid level of the questions becomes good. In addition to the validity test, another thing that needs to be done is measuring the level of reliability of the questions. So that the test questions given are consistent, reliable, measurable so that it makes the test questions very good. In addition, the test questions also need to be tested for differentiating power and the difficulty index of the questions so that the questions produced are getting better. Differentiation and difficulty index analysis will be carried out in other further studies. The following is a diagram of the validity of the test questions.
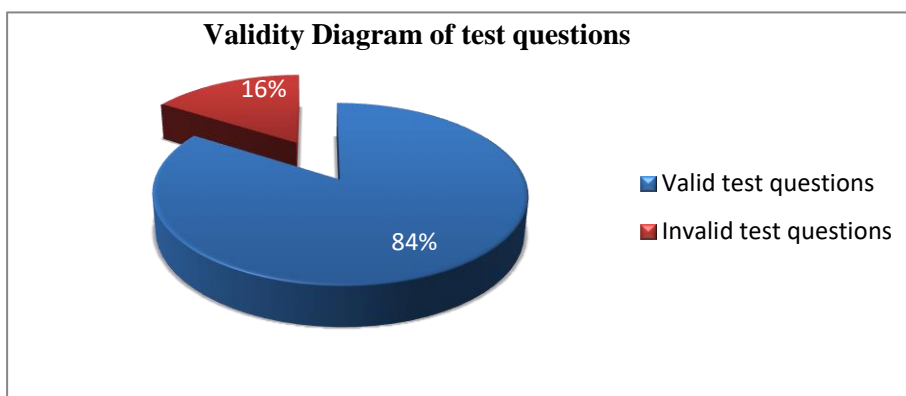

Figure 1.   Validity Diagram of test questions

The following is a table of test questions reliability criteria:

Table 2.   Reliability item test

Reliability Score Index	Reliability Criteria	Reliability Analysis Results
0,00 – 0,20	Very low	
0,21 – 0,40	Low	
0,41 – 0,70	Enough	0,8993
0,71 – 0,90	High	
0,91 – 1,00	Very High	

The results of the reliability analysis get a reliability value of 0.8993. The results of the reliability analysis values are then compared with the values with the reliability criteria. The conclusion from the results of the data reliability analysis is the level of reliability of the test questions with high quality. So that the test questions given can be trusted, consistent and stable anytime, anywhere and by whom the test is carried out, checked, and assessed. The test questions given are categorized as worthy to be used as test questions. So that the test questions can be used as one of the questions that are able to show the quality of student learning outcomes.

Analysis of the difficulty index of the test questions

The results of the analysis of the difficulty index of the test questions. The results of the analysis of the difficulty index of the test items are depicted in the form of the diagram below
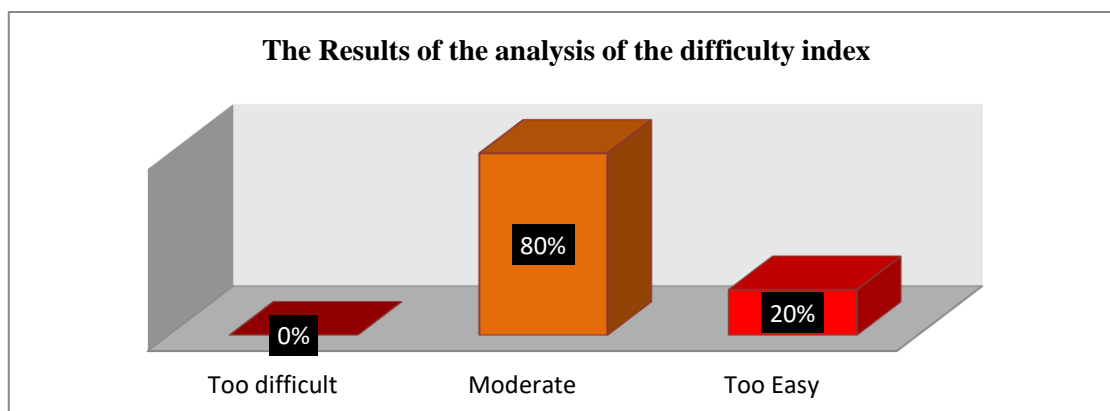

Figure 2.   The diagram of the results of the analysis of the difficulty index of test items

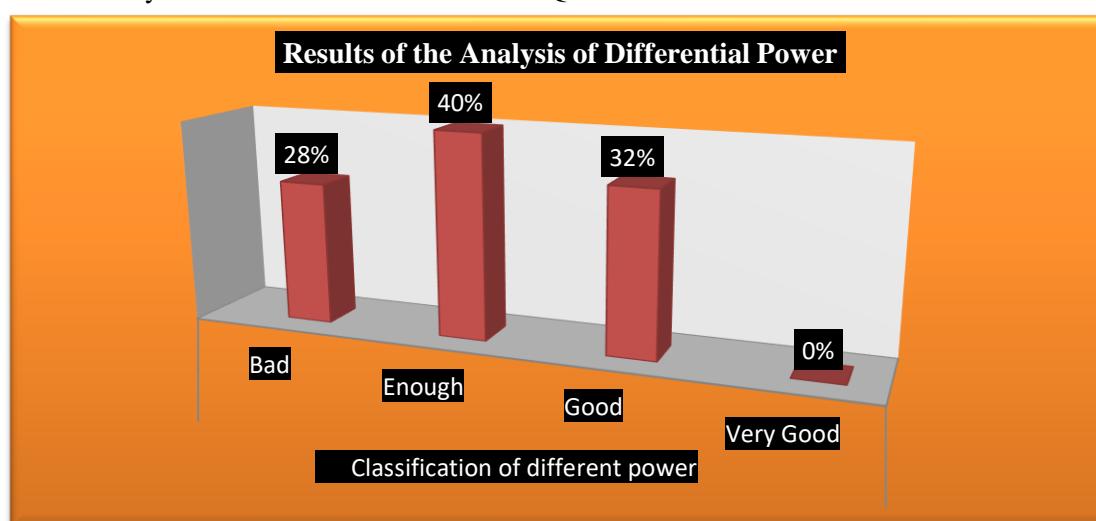Results of the Analysis of Differential Power of Test Questions

Figure 3.   Diagram of the results of the different power analysis of test questions

26

The results of the analysis of the difference in power of 25 test questions were given, the test questions with the classification of poor discrepancy were 7 questions with a percentage of 28%. Classification of the different power of test items is sufficient as much as 10 test questions with a percentage of 40%. Classification of different power of good test questions as many as 8 test questions with a percentage of 32%. The category of good/decent test items is a category of test items with sufficient and good classification. The results of the analysis show that the test questions are good and feasible at 72% of the total questions. Poor classification test questions need to be revised, repaired or if necessary replaced with other test questions. This is intended so that the test questions given will be good.

4. Conclusion

The results of the data analysis carried out showed that the level of validity of the test items was categorized as very high. Of the total 25 test questions given, there are 21 questions with a percentage of 84% of test questions categorized as valid. While the invalid questions are 4 questions with a percentage of 16% invalid questions. Invalid questions need to be corrected and revised or if the test questions are replaced. This is so that the quality of the test questions becomes a total valid question. The question reliability index is already high with a value of 0.8993. This shows that the test questions are reliable, consistent and stable whenever and wherever the test questions are given to students. Further analysis of test questions is needed in order to obtain test questions that are indeed able to measure and produce optimal student learning outcomes. The difficulty index analysis got 20 questions with moderate question criteria. This shows that 80% of the test questions given are included in the good/decent category. The test questions that are too easy are 20% of the total questions. Category test questions that are too easy need to be revised, corrected or if necessary replaced with other test questions. This is so that the test questions given will be good/worthy. The next analysis required is in the form of an analysis of the discriminatory power of test items that measure the level of ability in distinguishing high-ability students and low-ability students. The results of the analysis of the difference in power of 25 test questions given gave results in the form of test questions with a poor classification of 7 questions with a percentage of 28%. Classification of the different power of test items is sufficient as much as 10 test questions with a percentage of 40%. Classification of different power of good test questions as many as 8 test questions with a percentage of 32%. The category of good/decent test items is a category of test items with sufficient and good classification. Poor classification test questions need to be revised, repaired or if necessary replaced with other test questions. This is intended so that the test questions given will be good.

5. Acknowledgments

References

[1]   Saputra HD, Setiawan D, Amin B, et al. The effect of e-learning media on the improvement of learning outcomes in the Vehicle Body Construction course for students of the Department Automotive Faculty of Engineering UNP. 2021; 21: 7.

[2]   Saputra HD, Setiawan D, Yuvenda D, et al. PENINGKATAN KOMPETENSI SISWA MELALUI PELATIHAN TEKNOLOGI SEPEDA MOTOR. Suluah Bendang J Ilm Pengabdi Kpd Masy 2019; 19: 127.

[3]   Saputra HD, Ismet F, Andrizal A. Pengaruh Motivasi Terhadap Hasil Belajar Siswa SMK. INVOTEK J Inov Vokasional Dan Teknol 2018; 18: 25–30.

[4]   Saputra HD, Nasrun N, Wakhinuddin W. Development of Web-Based Learning Media in Vocational Secondary School. VOLT J Ilm Pendidik Tek Elektro 2018; 3: 37.

[5]   Klufa J. Multiple Choice Question Tests – Advantages and Disadvantages. 4.

[6]   Kizlik DB. Measurement, Assessment, and Evaluation in Education. 44.

[7]   Heo D, Anwar S, Menekse M. The Relationship Between Engineering Students' Achievement Goals, Reflection Behaviors, and Learning Outcomes. 11.

[8]   Love B, Hodge A, Grandgenett N, et al. Student learning and perceptions in a flipped linear algebra course. Int J Math Educ Sci Technol 2014; 45: 317–324.

[9]   Heale R, Twycross A. Validity and reliability in quantitative studies. Evid Based Nurs 2015; 18: 66–67.

[10]   Multiple-true-false questions reveal more thoroughly the complexity of student thinking than multiple-choice questions: a Bayesian item response model comparison. Int J STEM Educ; 6. Epub ahead of print December 2019. DOI: 10.1186/s40594-019-0169-0.

[11]   Department of anatomy, Pandit Dindayal Upadhyay Government Medical College, Gujarat, India., Chauhan P, Chauhan GR, et al. RELATIONSHIP BETWEEN DIFFICULTY INDEX AND DISTRACTER EFFECTIVENESS IN SINGLE BEST-ANSWER STEM TYPE MULTIPLE CHOICE QUESTIONS. Int J Anat Res 2015; 3: 1607–1610.

[12]   Loka Son A. INSTRUMENTASI KEMAMPUAN PEMECAHAN MASALAH MATEMATIS: ANALISIS RELIABILITAS, VALIDITAS, TINGKAT KESUKARAN DAN DAYA BEDA BUTIR SOAL. Gema Wiralodra 2019; 10: 41–52.