PILLAR OF PHYSICS EDUCATION
Vol. 16 (2), 2023, page.141-154
ISSN (Print) : 2337-9618
ISSN (Online) : 2684-8880

Analysis Quality of Equation Items in Learning Evaluation for Physics Subject in High Schools of The Mentawai Island Regency

Fredy Pratama[1], Yenni Darvina[1], Akmam[1], Wahyuni Satria Dewi[1]

[1] *Department of Physics, Universitas Negeri Padang, Jl. Prof. Dr. Hamka Air Tawar Padang 25131, Indonesia*

Corresponding author. E-mail: ydarvina@fmipa.unp.ac.id

ABSTRACT

The process of evaluating learning outcomes is one thing that cannot be separated by a teacher to reveal the extent to which students receive and learning understanding given in class. To be able to produce good measurable values, the instrument used must also be good. A good instrument must have standards that have been tested as a whole, in this case the question sheets used as instruments in the evaluation process must be measurable and tested in consist of validity, reliability, level of difficulty, and diffrentitation. This research is a descriptive research with a qualitative approach by taking several samples of schools in the Mentawai Islands District. The analysis was conducted on the evaluation instrument or sheet used by teachers in schools in the form of 4 aspects analysis, namely validity, reliability, level of difficulty, and diffrentitation. The research purpose was to be able to find out to what extent the quality level of the instrument or evaluation sheet used by teachers in the sampled schools. The research results obtained showed that of the 4 samples of schools that were taken and examined, it shows that in terms of validity for the whole school it can be declared valid with a range of high to low categories. Meanwhile, for the reliability aspect, it was noted that 3 out of 4 schools had reliable instruments, while 1 school was classified as unreliable. For aspects by level of difficulty and item differentitation, it is still classified as very low because there are still less than 50% of the questions that are of good quality and have sufficient differentitation.

Keywords: *Analysis; Validity; Reliability; Difficulty Level; Differentitation*

I. INTRODUCTION

In a process of learning there are assessment activities that become a benchmark for teachers or educators in seeing the development of their students after the process of learning. Assessment is often equated with evaluation activities. Evaluation is one of the critical components and stages that the teacher must complete in order to assess the learning effectiveness that has occurred. The evaluation results can be used as material for teacher feedback (feedback) in improving and perfecting programs or process of learning es in the classroom [1]. Evaluation is a component of the process and, as a whole, cannot be isolated from the learning activities process. Because in essence the evaluation of learning outcomes carried out by teachers or educators is carried out to monitor the process, progress and continuous improvement of student learning outcomes while at school, this is also one of the goals contained in the Law. In addition, the process of evaluating learning outcomes aims to measure competency achievement and improve the learning process, as well as provide standards for making progress reports on student learning outcomes [2].

Evaluation process cannot be conducted without going through the measurement process first. One way that is commonly done in the evaluation process is to give tests to students. The test is a set of stimuli given to someone with the intention of getting answers that form the basis for scoring. The test form consists of objective tests and essays [3]. The test as an evaluation tool has a very important role in the assessment. The test must be of good quality in order to get results that are actually in accordance with the reflection of the condition of students. In

order to know the test item quality, it is necessary to analyze the test item quality. The test items quality analysis or instruments shown from the value of validity, reliability, level of difficulty and the distinguishing power of the items or test instruments.

Validity relates to "accuracy" with measuring instruments. With an instrument was valid will also yield valid data. Alternatively, if the data generated by an instrument is valid, then the instrument is also valid [3]. The term "valid" is very difficult to find a replacement. There are those who replace the term valid with "valid", so that validity is replaced with validity. There are also those who translate the term valid with the word "correct", although the term "accurate" cannot yet cover all the meanings implied in the word "valid " so that the term validity is replaced with accuracy. In general, there are two types of validity: logical validity and empirical validity [4]

The word reliability in Indonesian is taken from the word reliability in English, which comes from the word reliable which means trustworthy. A test instrument can be said to be trusted if it gives consistent results or is mocked (consistent) when it is tested many times. If students are given the same test that is on different times, then each student will remain in the same order (ranking) or ridiculed in the group [3]. Based on these various meanings, in the field of measurement there are various terms to refer to the term reliability, namely some of them use the terms consistency, regularity, certainty, stability, and reliability. A reliable instrument is not necessarily valid. Materials that are broken at the ends, when used many times will produce the same (reliable) data, but are always invalid . This is because the instrument (meter) is damaged. Instrument reliability is a condition for determining the instrument's validity. So, even if the instrument was valid and generally dependable, reliability testing is required [3].

The level of difficulty is an indication of how difficult the question is for the participant. The level of difficulty of the items is revealed by the percentage of examinees on the questions given. The opinion of other researchers suggests that's "the calculation of the level of difficulty of the question is a measurement of how much the degree of difficulty of the item is". Based on this explanation, it can be interpreted that the level of difficulty or difficulty is the ratio between the number of students who answered the questions correctly and the number of test takers. The greater the number of students who answered correctly, the item has a lower level of difficulty. level of difficulty is one of the characteristics regarding the classical test quality theory , this characteristic has a good value if the resulting level of difficulty is of moderate value. An item with low scores or too difficult will be unfair to the ability of each student to be tested. Because every student has different abilities. There are high and low ability. Because of that the items that have a moderate level of difficulty are the middle way in assessing students' abilities . [1]

So, the level of difficulty of the item is the research of the item which aims to reveal the degree of difficulty of the item, namely questions that are classified as easy, medium and difficult. A good question is one that is neither too difficult nor too easy. Questions that are too difficult will frustrate the testee and will not want to try again, whereas questions that are too easy will not stimulate the testee's thinking ability and will not provide positive motivation . [5]

In another statement there is an opinion that stated "distinguishing power analysis examines the items with The goal of knowing the ability of the questions is to separate student who are capable from students who are less capable ". differentitation looks for differences in student abilities, which students have high abilities and students who have low abilities. In contrast to the level of difficulty which must have a moderate index, this test of differentitation if the item has a positive or high degree, the better the item is to distinguish students in the upper and lower classes. Testing items that have good quality is that the items have significant differentitation, meaning that the number of students who answer correctly must be more than students who answer incorrectly. If the conditions have been met then the item has positive differentiation . [6]

So, differentiation is the research of items with the aim of knowing and Differentiating between students with high and poor talents. The computation of differentiation is a measurement of how well an item may distinguish students who have mastered the competency from students who have not/have less learned the competency based on specified criteria. The greater an item's coefficient of differentiation, the better it is at distinguishing between students who grasp the competency and students who do not master the competency.

Question analysis aims to identify good, bad, and bad questions. With problem analysis, information can be obtained about the ugliness of a problem and instructions for making improvements. Test quality analysis is a step that can be taken to show the level of test quality, both for the entire test and for individual test items.

According to previous research, entitled. "The Class X High School Physics Question Items quality analysis ". The results of the qualitative analysis based on the construct, material, and language as a whole are adequate, but there are a few questions that need to be addressed in terms of the construct, material, and language elements [7]

The degree of test quality studied from the item items can be identified through the item analysis. Item analysis is an activity to examine the question items in the test whether they meet the requirements as a quality test. [8] From the these items analysis, it can be identified which items are good and which are not good and which items can be included in the question bank, revised, or discarded. Item analysis can be calculated through several aspects, namely Validity, Reliability, differentiation and level of difficulty. The form of questions used by formal institutions in the Final Semester Examination is an objective test (multiple choice) and essay (description). Questions in the Final Semester Examination must be of good quality in order to be able to measure the abilities of students' outcomes of learning precisely and accurately. For this reason, the questions must be analyzed to reveal the questions quality. Item analysis aims to identify good, bad, and bad questions. With problem analysis, information can be obtained about the ugliness of a problem and instructions for making improvements [9]. Otherwise, the test items of outcomes of learning analysis can be done from three aspects, namely: (1) in terms of the degree of difficulty of the items, (2) in terms of the differentiability of the items, (3) in terms of the distractor function [10].

However, in reality some educators pay less attention to the questions made quality. There are still many teachers who have not done the questions analysis they make because they think that doing the questions takes a long time and drains a lot of energy analysis. As a result, many of the items used in the test cannot produce correct data about student outcomes of learning. If a decision is made based on inaccurate data, then the decision cannot be justified. Therefore this research was conducted to see how far the existing test items quality would be analyzed by looking at the aspects of validity, reliability, differentiation and level of difficulty of the items to be analyzed.

II. METHOD

This research is a quantitative descriptive analysis in which the researcher describes and analyzes the data in the form of item test results for evaluation of the second semester of class XI high school physics in the Mentawai Islands Regency. This research uses a quantitative design because the information is embodied in the form of numbers. In this research, the information and research data obtained were in the form of quantitative data. Furthermore, the information and data were processed and analyzed using the Microsoft Office Excel program in order to obtain results that can be used to describe the items quality for evaluation in class XI semester II high school physics subjects in the Mentawai Islands Regency, in terms of Validity, Reliability, level of difficulty, and differentiation.

Data analysis techniques are one of the most important things in a research process, because with this technique the data obtained is tested and assessed so as to produce the desired analysis results. In this research, the analysis technique used by the researcher was to calculate the aspects that were used as a reference for assessing the question item quality, namely, validity, reliability, level of difficulty and distinguishing power.

The data used were obtained from 4 high schools in the Mentawai Islands Regency , namely: Senior High School 1 North South Pagai, Senior High School 1 North Pagai, Senior High School 1 South Siberut, and Senior High School 1 South Pagai. The data taken was in the form of even semester exam question sheets for physics for class XI as well as the evaluations results that had been conducted by students. Then the data was analyzed with the help of the MS Office Excel program. The analysis technique for the data obtained is to use the equation below.

Techniques for testing the validity of outcomes of learning test items can use the following equation [9]:

$$r_{pbi} = \frac{\mathrm{Mp}-\mathrm{Mt}}{\mathrm{SDt}}\sqrt{\frac{p}{q}} \qquad (1)$$

The following interpretation criteria to interpret the magnitude of the connection [9] :

Table 1. Distribution of Question Items Based On Validity

Validity Coefficient	Category
0.80 – 1.00	Very High
0.60 – 0.80	High
0.40 – 0.60	5 Enough
0.20 – 0.40	Low
0.00 – 0.20	Very Low

(Source: Ref [8])

Next reliability analysis can be revealed using the Kuder-Richardson formula (KR-21) namely [11] :

$$r_{11} = \left[\frac{n}{(n-1)}\right]\left[1 - \frac{M(n-M)}{nS^2}\right] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

Furthermore, to interpret the reliability coefficient of the test (r_{11}) Arikunto provides the following interpretation criteria [11]:

Table 2. Distribution of Question Items Based On Reliability

Limitation	Category
$0.80 \leq r_{11} < 1.00$	Very High
$0.60 \leq r_{11} < 0.80$	High
$0.40 \leq r_{11} < 0.60$	5 Enough
$0.20 \leq r_{11} < 0.40$	Low
$0.00 \leq r_{11} < 0.20$	Very Low

(Source: Ref [11])

To be able to calculate the index of level of difficulty of a question that has been tested we can use equation [1]:

$$P = \frac{B}{JS} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3)}$$

Furthermore, to interpret the calculations results at the level of difficulty of the questions, you can use the following interpretation [1]:

Table 3. Distribution of Question Items Based on Level of Difficulty

Limitation	Category
0.71 - 1.00	Easy
0.31 – 0.70	Currently
0.00 – 0.30	Hard

(Source: Ref [11])

To be able to calculate the differentiation of each item we can use the following equation [1] :

$$D = \frac{BA}{JA} - \frac{BB}{JB} = Pa - Pb \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(4)}$$

Calculation of the differentiation index can be concluded with the following reference [9] :

Table 4. Distribution of Question Items Based On Differentiation

Differentiation Index	Criteria
0.00 – 0.20	Bad *(poor)*
0.21 – 0.40	Enough *(satisfactory)*
0.41 – 0.70	Good
0.71 – 1.00	Very Good

(Source: Ref [11])

The criteria used to reveal the quality level of the item items were adapted from the Scale of Likert as follows [8] :

Table 5. Distribution of Question Items Based On Question Items Quality

Number of Criteria Met (Validity, Level of Difficulty, Differentiation, and Reliability)	Question Items Quality	Revision	Enter the Question Bank
4	Very Good	There isn't any	Yes
3	Good	Minor	Not Yet
2	Currently	Currently	Not Yet
1	Not Good	Discarded	No
1	Very Not Good	Discarded	No

(Source: Ref [8])

Explanation from the table of quality criteria for the items above

a. Items are very good if they meet four good criteria, namely Validity, level of difficulty, and differentiation. In this condition, the item can be entered into the question bank.
b. Items are good if they meet three of the four criteria for good questions (Validity, level of difficulty, and differentiation). In this condition, the item questions cannot be entered into the question bank. The questions must be revised so that they meet the four criteria.
c. An item is moderate if it meets two of the four criteria for a good item (Validity, level of difficulty, and differentiation). In this condition, the items cannot be entered into the question bank. The questions must be revised so that they meet the four criteria.
d. Items are not good if they meet one of the four criteria for good questions (Validity, level of difficulty, and differentiation). In this condition, the item cannot be entered into the question bank. The questions require significant revision so it's better not to use them.
e. Items are not very good if they do not meet all the criteria for good questions (Validity, level of difficulty, and differentiation). In this condition, the item cannot be entered into the question bank. The questions require significant revision so it's better not to use them.

III. RESULTS AND DISCUSSION

Existing data is processed using the help of the Microsoft Office Excel application. The data obtained is processed according to the aspects or criteria that are used to measure the item quality, which includes validity, reliability, level of difficulty and distinguishing power. From each of the existing aspects, the researcher searched for and divided the data into each school that was the research sample so that the data displayed was the processing the data result contained in each school used as a sample.

After processing the data, the values or criteria are obtained from the aspects that are used to measure the item quality, which the researcher then groups back based on a scale that is in accordance with the theory so that it shown how the item quality is on the instrument used. From the results that have been obtained, it shows that most of the instruments analyzed and used as samples still have quality in the lower middle category or in the moderate to poor category. This is caused by several factors, including too many invalid questions so that the reliability value of the instrument is also low or unreliable. In addition, the level of difficulty and differentiation of the existing item items are still in the easy category, even though at the middle school level and above the item items used should be at C4 to C6 levels with varying levels of difficulty so that students are not too easy in answer the questions given. besides that there was no process of analyzing the questions conducted by the teacher before using the questions on students also became one of the things that influenced the research results.

The data analysis results that has been conducted to see the even semester end exam questions quality for class XI in high school physics subjects in Mentawai Islands Regency are seen from the aspects of validity, reliability, distinguishing power and level of difficulty shown below.

A. Validity of Question Items

The first data taken came from Senior High School 1 North South Pagai school, the total number of class XI students who took the even semester final exam for the 2021/2022 academic year, namely 102 students, with a correlated variable of 2. Thus N = X-2 = AX, while the significance level was obtained from the rtable of 0.19. if so, the provisions of the item can be said to be valid if it is at rcount ≥ 0.19. Based on the analysis results of 25 multiple-choice items on the Even Semester End Examination questions for class XI in the 2021 academic year at Senior High School 1 North South Pagai which can be declared valid are 20 questions (80%) and questions which are declared invalid are 5 questions (20%)).

The second data taken comes from Senior High School 1 North Pagai school, the total number of class XI students who are taking the even semester final exam for the 2021/2022 school year is 128 students with a correlated variable of 2. Thus N = X-2 = AX, whereas the significance level was obtained from the rtable of 0.17. if so then the provisions of the item can be said to be valid if it is at rcount ≥0.17. As for the even semester end exam questions for class XI high school physics subjects for the 2021/2022 school year at Senior High School 1 North Pagai, Mentawai Islands Regency, there were 12 questions (60%) that could be declared valid and 8 questions (40%) that were declared invalid. .

The third data taken came from Senior High School 1 South Siberut school, the total number of class XI students who took the even semester final exam for the 2021/2022 academic year, namely 82 students, with a correlated variable of 2. Thus N = X-2 = AX, whereas the significance level was obtained from the rtable of 0.21. if so, the provisions of the item can be said to be valid if it is at rcount ≥ 0.21. Based on the analysis results of 25 multiple-choice items on the Even Semester End Examination questions for class XI in the 2021 academic year at Senior High School 1 South Siberut, there are 11 questions (44%) that can be declared valid and 14 questions that are declared invalid (56%).

The last sample is Senior High School 1 South Pagai school, the total number of class XI students taking the even semester final exam for the 2021/2022 school year is 47 students with the correlated variable being 2. Thus N = X-2 = AX, while the significance level is get from rtable of 0.28. if so, the provisions of the item can be said to be valid if it is at rcount ≥ 0.28. Based on the analysis results of 25 multiple-choice items on the Even Semester End Examination questions for class XI in the 2021 academic year at SMA 1 Pagai Urata Selatan which can be declared valid are 11 questions (44%) and 14 questions which are declared invalid (56%) . As for the data from the validity aspects of the four schools analysis that were sampled in the research, they can be shown in the table below.

Table 6. Distribution of Question Items Based On Question Items Quality

No	School Name	Validity Index	Question Items	Amount	Presentation (%)
1	Senior High School 1 North South Pagai	≥ 0.19 (Valid)	1,2,3,5,6,8,10,11,12,13, 14,15,16,18,20,21,22,23, 24,25	20	80%
		< 0.19 (Invalid)	4,7,9,17,19	5	20%
2	Senior High School 1 North Pagai	≥ 0.17 (Valid)	1-6,10-14,18-24	17	68%
		< 0.17 (Invalid)	3,7,8,9,15,16,17,25	8	32%
3	Senior High School 1 South Siberut	≥ 0.21 (Valid)	1,5,6,13,14,15,16,20,21, 22,23	11	44%
		< 0.21 (Invalid)	2,3,4,7-12,17,18,19,24,25	14	56%
4	Senior High School 1 South Pagai	≥ 0.28 (Valid)	2,3,9,14,15,16,19,20,21, 22,24	11	44%
		< 0.28 (Invalid)	1,4,5,6,7,8,10-13,17,18,23,25	14	56%

(Source: Reseacrh Data)

The results for the final semester exam question sheets for class XI semester II in several schools that were taken as samples, namely, Senior High School 1 North South Pagai, Senior High School 1 North Pagai, Senior High School 1 South Siberut, Senior High School 1 South Pagai showed that the instrument as a whole had items that were valid however, is in the middle to lower category. These results show that an instrument that can be considered cannot be used to accurately measure students' abilities because validity is one aspect that is expected to be able to measure precisely and accurately the abilities of students [4]. In addition, validity can also be seen whether a test or measuring instrument is valid or not depending on the extent to which the test items reflect (present) the aspects to be measured . So it is hoped that the items made do not deviate from the aspects of the variables to be measured. Validity also has an important role in achievement tests, by providing a grid that includes the content and competencies to be measured [12]. The low level of validity of the instrument being analyzed is also caused by the habits of students who are not used to solving problems in questions at level C4 and above, while students are generally more used to working on questions in the range of levels C1 - C3.

B. Reliability

The first sample is reliability testing for even semester final exam questions for class XI high school physics subjects for 2021/2022 at Senior High School 1 North South Pagai , Mentawai Islands Regency. The researcher was assisted by the MS office Excel program which was based on a value of $r11 > 0.19$, so the items tested had high reliability, but vice versa if the value of $r11 \leq 0.19$, the items tested had low reliability or could be said to be unreliable. From the calculations results that have been done by the researcher, the even semester final exam questions for class XI high school physics for the 2021/2022 school year at SMA N 1 Pagai Utara Selatan, Mentawai Islands Regency have a reliability of 0.60 so we got conclussion that the items are declared reliable but category in medium.

The next data shows the reliability analysis results of the question instrument used at Senior High School 1 North Pagai shows that the value of $r11 = 0.17$. therefore if $r11 > 0.17$ then the item being tested has high reliability, but vice versa if the value of $r11 \leq 0.17$ then the item being tested has low reliability or can be said to be unreliable. The calculations results that have been conducted by the researcher, the even semester final exam questions for class XI high school physics for the 2021/2022 academic year at SMA N 1 Pagai Utara, Mentawai Islands Regency have a reliability of 0.16 so we got conclussion that these items are stated to be unreliable.

Next, the reliability analysis results of the question instrument used at Senior High School 1 South Siberut show that the value of $r11 = 0.21$. therefore if $r11 > 0.21$ then the item being tested has high reliability, but vice versa if the value of $r11 \leq 0.21$ then the item being tested has low reliability or can be said to be unreliable. From the calculations results that have been conducted by the researcher, the even semester final exam questions for class XI high school physics for the 2021/2022 academic year at SMA N 1 Siberut Selatan, Mentawai Islands Regency have a reliability of 0.24 so we got conclussion that the item is stated to be reliable but with category in medium.

The last sample for the reliability analysis results of the question instrument used at Senior High School 1 South Pagai shows that the value of $r11 = 0.28$. therefore if $r11 > 0.28$ then the item being tested has high reliability, but vice versa if the value of $r11 \leq 0.28$ then the item being tested has low reliability or can be said to be unreliable. From the calculations results that have been done by the researcher, the even semester final exam questions for class XI high school physics for the 2021/2022 academic year at SMA N 1 Pagai Selatan, Mentawai Islands Regency have a reliability of 0.33 so we got conclussion that the items are declared reliable but with category in medium.The analysis results of the reliability aspect of the four schools that were sampled in the research are shown as a whole in the table below :

Table 7. Distribution of Question Items Based On Question Items Quality

No	School Name	Category	Mark	Conclusion
1	Senior High School 1 North South Pagai	Mean Total Score	18.99019608	Reliable
		Standard Deviation	3,282	
		Reliability Coefficient	0.6	
		r Table	0.1946042	

2	Senior High School 1 North Pagai	Mean Total Score	17.976563	Not Reliable
		Standard Deviation	2,448	
		Reliability Coefficient	0.16	
		r Table	0.1736231	
3	Senior High School 1 South Siberut	Mean Total Score	17.97560976	Reliable
		Standard Deviation	2,566	
		Reliability Coefficient	0.24	
		r Table	0.217185	
4	Senior High School 1 South Pagai	Mean Total Score	20.31914894	Reliable
		Standard Deviation	2,353	
		Reliability Coefficient	0.33	
		r Table	0.287563	

(Source: Reseacrh Data)

These results are divided into two categories, some are in accordance with the theory which states that if one of the requirements for a good question is that it has been tested for reliability, some are not in accordance with the theoretical research which states that one of the requirements for a good item and can be used as an evaluation instrument is questions with high reliability. A reliable instrument is not necessarily valid. Materials that are broken at the ends, when used many times will produce the same (reliable) data, but are always invalid . This is because the instrument (meter) is damaged. Instrument reliability is a requirement for testing the validity of the instrument. So, even though instrument was valid are generally reliable, reliability testing needs to be done [13]. The causes of questions with low reliability generally occur due to the lack of valid item items listed in the question instrument made by the teacher. So that the existing questions become less reliable. As for fixing it, it can be done by adding valid items, because the more valid items in an evaluation instrument, the higher the reliability value of the instrument will be.

C. Level of Difficulty

Analysis for the level of difficulty of the items in this research was conducted using a index of difficulty whose results were then interpreted into three criteria or categories namely; questions with a index of difficulty of 0.00 to 0.30 are questions in the difficult category; questions with a index of difficulty of 0.31 to 0.70 are questions that fall into the category in medium; questions with a index of difficulty of 0.71 to 1.00 are questions in the easy category. Based on the research results on even semester final exam questions for Class XI physics for the 2021/2022 academic year at Senior High School 1 North South Pagai, Mentawai Islands Regency which has been conducted on 25 multiple choice questions, it shown that 20 questions are included in the easy question category. (80%), questions that fall into the category of moderate questions amount to 5 questions (20%).

Next, for the research results on even semester final exam questions for Class XI physics for the 2021/2022 school year at Senior High School 1 North Pagai, Mentawai Islands Regency, which has been conducted on 25 multiple choice questions, it shown that 18 questions are included in the easy question category. (72%), questions that fall into the category of moderate questions amount to 7 questions (28%). Furthermore, the research results on the even semester final exam questions for Class XI physics for the 2021/2022 academic year at Senior High School 1 South Siberut. Mentawai Islands Regency, which has conducted 25 multiple choice

items, it shown that 18 questions (72%) are included in the easy question category, 7 questions (28%) are included in the moderate item category.

The next sample is the research on even semester final exam questions result for Class XI physics for the 2021/2022 academic year at Senior High School 1 South Pagai. Mentawai Islands Regency, which has done 25 multiple choice items, it shown that 21 questions (84%) are included in the easy item category, 4 questions (16%) are included in the moderate item category. The analysis results of the level of difficulty aspects of the four schools that were sampled in the research shown in the table below.

Table 8. Distribution of Question Items Based on level of difficulty

No	School Name	Category	Question Items	Amount	Presentation (%)
1	Senior High School 1 North South Pagai	0.71 – 1.00 (Easy)	1,2,3,4,5,6,7,8,9,10,11,13, 14,15,16,17,18,19,20,21	20	80%
		0.31 – 0.70 (Medium)	12,22,23,24,25	5	20%
2	Senior High School 1 North Pagai	0.71 – 1.00 (Easy)	1-12, 15, 18, 20, 21, 22, 25	18	72%
		0.31 – 0.70 (Medium)	13, 14, 16, 17, 19, 23, 24	7	28%
3	Senior High School 1 South Siberut	0.71 – 1.00 (Easy)	1-10, 14,15,16,18,21,22,23,25	18	72%
		0.31 – 0.70 (Medium)	11,12,13,17,19,20,24	7	28%
4	Senior High School 1 South Pagai	0.71 – 1.00 (Easy)	1-8,10,12,13,14,16, 20,22,23,24,25	21	84%
		0.31 – 0.70 (Medium)	9,11,15,21	4	16%

(Source: Reseacrh Data)

The level of difficulty is an indication of how difficult the question is for the participant. The difficulty level of the items is revealed by the percentage of examinees on the questions given. Zainal Arifin put forward "the calculation of the level of difficulty of the question is a measurement of how much the degree of difficulty of the item is". Based on this explanation, it can be interpreted that the level of difficulty or difficulty is the ratio between the number of students who answered the questions correctly and the number of test takers. The greater the number of students who answered correctly, the item has a lower level of difficulty [1]. Difficulty level is one of the characteristics regarding the classical test quality theory, this characteristic has a good value if the resulting level of difficulty is of moderate value. An item with low scores or too difficult will be unfair to the ability of each student to be tested. Because every student has different abilities. There are high and low ability.

The degree of difficulty or level of difficulty possessed by each item is the first indicator of item quality. These items are considered good if they are neither too difficult nor too easy, or if the degree of difficulty of the item is reasonable or sufficient. The number that indicates the difficulty and ease of a question is called the index of difficulty (difficuly index). The index of difficulty is between 0.00 and 1.00. This index of difficulty shows the level of difficulty of the question. Questions with a index of difficulty of 0.00 indicate that the item is too difficult, otherwise 1.00 indicates that the item is too easy [14]

So, the level of difficulty of the item is the research of the item which aims to reveal the degree of difficulty of the item, namely questions that are classified as easy, medium and difficult. A good question is one that is neither too difficult nor too easy. Questions that are too difficult will frustrate the testee and will not want to try again, whereas questions that are too easy will not stimulate the testee's thinking ability and will not provide positive motivation [14].

From the analysis results that has been conducted, the results show that for the level of difficulty quality the instrument being analyzed it is still in the easy and category in medium, there are still no questions with a level of difficulty in the difficult category in the instrument being analyzed. Thus it can be concluded based on the existing theory that the instrument being analyzed is still classified as an easy type of problem.

D. Differentitation

At this stage of the differentitation items analysis, it is the same as the level of difficulty items analysis. Researchers used a differentitation index which was then divided into 4 categories namely; if the value of D = 0.00 – 0.20 means that the differentitation in the questions is in the bad category; if D = 0.21 – 0.40 it means that the differentitation in the questions is in the sufficient category; if D = 0.41 – 0.70 it means that the differentitation in the questions is in the category in good; whereas if the value of D = 0.71 – 1.00 means that the value of the differentitation in the question is in the category in very good and if the value of D = Negative then the value of the differentitation of the item is not very good and it is recommended to be replaced/thrown away.

The research on the even semester final exam questions results in class XI high school physics for the 2021/2022 academic year at Senior High School 1 North South Pagai Regency, Mentawai Islands Regency, have been conducted on 25 multiple choice questions which have differentiating criteria. It shown that as many as 24 questions (96%) items are categorized as bad, 1 item (4%) items are considered sufficient.

Furthermore, the research on even semester final exam questions results in class XI high school physics subject for the 2021/2022 academic year at Senior High School 1 North Pagai, Mentawai Islands Regency, have been conducted on 25 multiple choice questions that have differentiating criteria. It shown that as many as 21 questions (84%) of the items were categorized as bad, 4 items (16%) were considered sufficient.

Next, the research on even semester final exam questions results in class XI high school physics for the 2021/2022 school year at Senior High School 1 South Siberut, Mentawai Islands Regency, which have been conducted on 25 multiple choice questions that have differentiating criteria. It shown that as many as 21 questions (88%) items were categorized as bad, 4 questions (16%) items were considered sufficient.

The last sample analyzed, the research on even semester final exam questions results in class XI high school physics for the 2021/2022 academic year at Senior High School 1 South Siberut, Mentawai Islands Regency, were conducted on 25 multiple choice items that had differentiating criteria. It shown that all questions have a differentitation value that belongs to the bad category. The differentitation analysis results that has been conducted on the instruments collected show that the instruments obtained still do not have good differentitation. The differentiating power aspects analysis results of the four sample schools shown in the following table.

Table 9. Distribution of Question Items Based on Differentiation Power

No	School Name	Category	Question Items	Amount	Presentation (%)
1	Senior High School 1 North South Pagai	0.21 – 0.40 (Enough)	3	1	4%
		0.00 – 0.20 (Poor)	1,2,4-25	24	96%
2	Senior High School 1 North Pagai	0.21 – 0.40 (Enough)	2, 8, 14, 24	4	16%
		0.00 – 0.20 (Poor)	1,3,4-6,7,9-13,15-18,20-21, 25	21	84%
3	Senior High School 1 South Siberut	0.21 – 0.40 (Enough)	2,8,19	3	12%
		0.00 – 0.20 (Poor)	1,3-7,9-18,20-25	22	88%
4	Senior High School 1 South Pagai	0.21 – 0.40 (Enough)	-	-	-
		0.00 – 0.20 (Poor)	1-25	25	100%

(Source: Reseacrh Data)

Distinguishing power analysis examines the items with the aim of knowing the ability of the questions to distinguish students who are classified as capable from students who are classified as less. differentitation looks for differences in student abilities, which students have high abilities and students who have low abilities. In contrast to the level of difficulty which must have a moderate index, this test of differentitation if the item has a positive or high degree, the better the item is to distinguish students in the upper and lower classes. Testing items that have good quality is that the items have significant differentitation, meaning that the number of students who answer correctly must be more than students who answer incorrectly. If the conditions have occurred, then the item already has positive differentitation [6].

Differentiation can also be applied to the items in order to identify and distinguish between pupils with high and low talents. Differentiation calculation is a measurement of the amount to which an item can separate students who have mastered the competency from students who have not/have less learned the competency based on specified criteria. The higher an item's coefficient of differentiation, the better it is at distinguishing between students who master the competency and students who do not[1].

According to the text, differentitation relates to the degree of item ability to distinguish well the behavior of test takers in the developed test. Questions can be said to have differentitation if these questions can be answered by students with low abilities. If a question can be answered by smart or poor students, it means that the question does not have differentitation, likewise if the question cannot be answered by smart students and students who are lacking, it means that the question is not good because it does not have differentitation [15] . This has not been found in the instrument questions results analysis that are owned, so it is still not in the category of having good differentitation or according to good standards.

E. Question Items Quality

The items quality analysis was conducted by drawing conclusions using the aspects previously described, namely validity, level of difficulty, distinguishing power and reliability of the instrument itself. The criteria used to facilitate the interpretation of the items quality were adapted from the Scale of Likert , namely: very good, good, good enough, not good, and very bad.

Based on the item the Even Semester Final Examination results analysis for Physics Class XI Academic Year 20 20 at Senior High School 1 North South Pagai, Mentawai Islands Regency as a whole the multiple choice questions that have been described above, it shown that the items that meet all the criteria are Validity, level of difficulty, differentitation, is a question of good quality. Items with quality in very good totaled 0 items or equal to (0 %), good quality items totaled 2 items or equal to (8 %), items with sufficient quality totaled 7 items or equal to (28 %) and item the bad quality totaled 16 items or equal to (64 %). Items that are of poor quality should not be used and replaced with new questions .

Based on the Even Semester Final Examination Questions results analysis for Class XI Physics for the 2021/2022 Academic Year at Senior High School 1 North Pagai, Mentawai Islands Regency as a whole the multiple choice question items that have been described above, it can be seen that the item meets all the criteria, namely Validity , level of difficulty, differentitation, are questions that are classified as of good quality. Items with quality in very good totaled 0 items or equal to (0%), good quality items totaled 2 items or equal to (8%), items with sufficient quality totaled 5 items or equal to (20%) and items the bad quality totaled 18 items or by (72%). Items that are of poor quality should be discarded and replaced with new items.

Based on the item Even Semester Final Examination Questions results analysis for Class XI Physics for the 2020 Academic Year at Senior High School 1 South Siberut, Mentawai Islands Regency as a whole the multiple choice question items that have been described above, it can be seen that the item meets all the criteria, namely Validity, Level Difficulty, differentitation, is a question that is classified as of poor quality. Items with quality in very good totaled 0 items or equal to (0%), good quality items totaled 0 items or equal to (0%), items with sufficient quality totaled 4 items or equal to (16%) and items the bad quality totaled 21 items or by (84%). Items that are of poor quality should not be used and replaced with new questions.

Based on the item Even Semester Final Examination Questions results analysis for Class XI Physics for the 2020 Academic Year at Senior High School 1 South Siberut, Mentawai Islands Regency as a whole the multiple choice question items that have been described above, it can be seen that the item meets all the criteria, namely Validity, Level Difficulty, differentitation, is a question that is classified as of poor quality. Items with quality in very good totaled 0 items or equal to (0%), good quality items totaled 0 items or equal to (0%), items with sufficient quality totaled 2 items or equal to (8%) and item the bad quality totaled 23 items or by (92%). Items that are of poor quality should not be used and replaced with new questions.

In another study that was carried out with the title "Analysis of odd semester summative test items for physics subjects" the conclusions drawn from the results of this study indicate that the quality of the odd semester final exam items for physics class X IPA SMA Negeri 1 Remboken 2019/2020 academic year in terms of validity, reliability, discriminating power, and level of difficulty for description questions there is 1 item that has good quality, 3 questions that have moderate quality need to be revised so that they can be reused and 1 question number that has bad quality can be discarded and can't be reused. The quality of the odd semester end exam questions for physics class X IPA SMA Negeri 2 Tondano for the 2019/2020 academic year in terms of validity, reliability, discriminating power, level of difficulty and effectiveness of the distractor for multiple choice questions, there are 8 question numbers that are of good quality, 7 question numbers have moderate quality that need to be repaired so they can be reused and 10 question numbers have bad and very bad quality so they can be discarded and cannot be reused. [16].

In other research that is still related to the analysis of student evaluation test instruments in physics with the title "Analysis of the quality of class X senior high school physics questions" obtained quantitative analysis results based on difficulty level, the proportion of difficulty levels was not balanced. The ability to discriminate questions is not functioning properly. Most of the effectiveness of the distractor in the Physics question bank in class X SMA Negeri 2 Bunut Hulu is functioning properly, so the effectiveness of the distractor can be accepted and used. As many as 65% of questions are valid. Reliability 0.65 is in the moderate category so that the questions can be trusted to evaluate students [17]. The use of assessment instruments must be in accordance with the rules and conditions that have been set. Thus we need an assessment instrument in the assessment system to determine the level of success of the learning process that is able to measure all aspects of student competency. Therefore the assessment instrument is important and must be implemented [18].

The following is a distribution table of the even semester exam questions results analysis for class XI physics for the 2021/2022 academic year for the four schools that were sampled in the research in terms of the aspects and criteria that have been described.

Table 10. Distribution of Question Items Based On Question Items Quality

No	School Name	Questions Items Quality	Question Items	Amount	Presentation (%)
1	Senior High School 1 North South Pagai	Very good	-	-	0%
		Good	22,24	2	8%
		Enough	3,6,12,14,15,23,25	7	28%
		Not good	1,2,4,5,7-11,13,16-18,19,20,21	16	64%
		Very Not Good	-	-	0%
2	Senior High School 1 North Pagai	Very good	-	-	0%
		Good	14, 24	2	8%
		Enough	2, 10, 13, 19, 20	5	20%
		Not good	1, 3-9, 11,12, 15-18, 21-23, 25	18	72%
3	Senior High School 1 South Siberut	Very Not Good	-	-	0%
		Very good	-	-	0%
		Good	-	-	0%
		Enough	13,19,20,23	4	16%
		Not good	1-12,14-18,21,22,24,25	21	84%
4	Senior High School 1 South Pagai	Very Not Good	-	-	0%
		Very good	-	-	0%

(Source: Reseacrh Data)

Based on the results described above, for the items quality analysis as a whole the instruments analyzed are still not included in good instruments and are worthy of meeting existing standards. This is because of the 4 instrument samples taken, all of them have items in the sufficient and not good scale ranges. Even though the instruments used should have been standardized and of good quality. As described in the introduction item quality analysis is a process done to reveal the degree of difficulty of a test or item, both the test as a whole and the items that make up the test. The exam is expected to explain a sample of behavior and generate an objective and accurate value for assessing learning results. If the teacher's test is inadequate, the outcomes will be inadequate as well. This might be harmful to pupils, as it implies that the results acquired by students are not objective. As a result, the teacher's test must be of higher quality in a variety of ways. Tests should be prepared in compliance with the test preparation principles and methods. It is vital to examine the test quality after use to reveal whether the test utilized is good or not good [1]. Another opinion suggests that "question analysis is conducted to find out whether a question is functioning or not :. From the above understanding we can apply that item analysis is the activity of analyzing each item in detail with a particular testing method. The purpose of this test is for the items to become a quality assessment tool [13].

Good test quality results from a series of processes that need to be analyzed for the quality of the test instruments made. So that the results of the assessment carried out meet the standards and meet the criteria and aspects being assessed. Therefore the process of quality analysis of this test instrument is very important to be carried out by teachers or teaching staff. Because the results of the analysis that the researchers carried out on the existing samples were still below the standard they should be [19]. And from the results of the analysis that the researchers have done on the existing samples, the results are still below the standard that should be.

The questions must be analyzed first in order to determine the quality of the questions. The question is said to be of good quality if it meets the characteristics of the assessment of the items said to be valid and reliable. The test given to students must be considered by the teacher. So far, in carrying out assessments, teachers only look at the final results obtained by students, but teachers must also evaluate the tests given to students [20]. Of course, the researcher hopes that this research can be a picture in the future so that the assessment instruments to be used can be analyzed in advance for the quality of the instruments.

IV. CONCLUSION

From the analysis results that was conducted, it was found that the evaluation instrument sheets used by teachers in the 4 sample schools had different results . For the aspect of validity, as a whole it can be said to be valid with a range of high – very low categories or at a value of 0.60 – 0.20. Meanwhile, for the reliability aspect, there were 3 out of 4 sample schools where the evaluation instrument was classified as reliable or the reliability coefficient value (r_{11}) was greater than the $r_{table\ value}$ and there was 1 school whose evaluation instrument was still not reliable because the reliability coefficient value (r_{11}) was higher greater than the value of r_{table} . Next, for the level of difficulty aspect of the instrument questions, it shows that the range of categories in the questions analyzed are mostly still in the easy category and a few are in the category in medium, an average of only about 25% of the total sample has a medium level of difficulty category, the rest are in the easy category. . This also occurs in the aspect of differentitation where the average of the entire sample still has a poor level of differentitation so that it is easy for students to guess, from the analysis obtained as a whole the average for differentitation of questions is around 10% of the total sample having the differentitation. sufficient category differentiator. So that for the general school samples quality aken for the instrument to be analyzed, they still have a quality that is classified as not good enough to be used as a tool or instrument for evaluating learning for students, before the teacher or educator gives an evaluation to students, the instrument used by the teacher must already be well tested and standardized so that the evaluation process is conducted right on target.

REFERENCES

[1] Arifin, Zainal, Evaluasi Pembelajaran. Bandung: PT Remaja Rosdakarya. 2016
[2] Septiana, Nurul. Analisis Butir Soal Ulangan Akhir Semester (UAS) Biologi Tahun 2015/2016 Kelas X dan XI Pada MAN Sampit. Edu Sains 4, no. 2 : h. 115. 2016
[3] Widoyoko, Eko Putro. Evaluasi Program Pembelajaran. Yogyakarta : Pustaka Pelajar. 2014
[4] Arikunto, Suharsimi. Dasar-Dasar Evaluasi Pendidikan. Jakarta : Bumi Aksara. 2016
[5] Syamsudduha, St. Penilaian Kelas. Makassar: Alauddin University Press. 2016
[6] Sudjana, N. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya. 2014
[7] Anita, A., Tyowati, S., & Zuldafrial, Zainal. Analisis kualitas butir soal fisika kelas x sekolah menengah atas. Edukasi: Jurnal Pendidikan, 16(1), 35-47.Wika Sevi Oktanin. Analisis Butir Soal Ujian Akhir Mata

Pelajaran Ekonomi Akuntansi. Jurnal Pendidikan Akuntansi Universitas Negeri Yogyakarta, Vol. XIii, No.1, Tahun 2018. 2018
[8] Arikunto, Suharsimi. Dasar-Dasar Evaluasi Pendidikan. Jakarta : Bumi Aksara. 2016
[9] Sudjono, Anas. Pengantar Evaluasi Pendidikan. Yogyakarta: Rajawali. 1995
[10] Arikunto, S. *Prosedur Penelitian Suatu Pendekatan Praktik.* Jakarta: PT Rineka Cipta. 2014
[11] Purwanto, Ngalim. Prinsip-Prinsip dan Teknik Evaluasi Pengajaran. Bandung: Remaja Rosdakarya.
[12] Harsi, Deradi. Analisis Kualitas Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Pemrograman WEB di SMK Kelas X Teknik Komputer Jaringan Kota Yogyakarta Tahun Ajaran 2015/2016 Skripsi (Yogyakarta : Fak. Tekhnik, 2016), h. 20. 2003
[13] Amirono, Daryanto. Evaluasi dan Penilaian Pembelajaran Kurikulum 2013. Yogyakarta: Gava Media. 2016
[14] Purwanto. Evaluasi Hasil Belajar. Yogyakarta: Pustaka Belajar. 2011
[15] D. Susanti, V. Fitriani, and L. Y. Sari, "Validity of module based on project based learning in media biology subject," *J. Phys. Conf. Ser.*, vol. 1521, no. 4, pp. 0–6, 2020, doi: 10.1088/1742-6596/1521/4/042012.
[16] Umacina, N. E. P., Mandolang, A. H., & Rondonuwu, I. T. Analisis butir soal sumatif semester ganjil mata pelajaran fisika. *Charm Sains: Jurnal Pendidikan Fisika*, *1*(2), 33-38. 2020.
[17] Anita, A., Tyowati, S., & Zuldafrial, Z. Analisis kualitas butir soal fisika kelas x sekolah menengah atas. *Edukasi: Jurnal Pendidikan*, *16*(1), 35-47. 2018.
[18] Maulidah, H., Sukarno, S., & Syefrinando, B. ANALISIS KUALITAS INSTRUMEN TES FISIKA KELAS X MENGGUNAKAN SOFTWARE ANATES. *Physics and Science Education Journal (PSEJ)*, 153-162. 2022.
[19] Asrul, dkk. Evaluasi Pembelajaran. Medan: Perdana Mulya Sarana. 2014.
[20] Novia, T., Wardani, A., Canda, C., Nurdi, N., & Nurmasyitah, N. Analisis Validitas dan Reliabilitas Butir Soal UTS Fisika Kelas X SMA Swasta Muhammadiyah 4 Langsa. *GRAVITASI: Jurnal Pendidikan Fisika dan Sains*, *3*(01), 19-22. 2020.