ISSN 2987-4270
PLE
Physics Learning and Education
Vol. 3 No. 4 (2025)

Analysis of Students' Critical Thinking Ability Instruments in Thermodynamics Material Phase F SMA/MA Reviewed from Validity, Reliability, Level of Difficulty and Differentiation

Nurjannah[1*], Emiliannur[1]

[1] *Department of Physics, Universitas Negeri Padang, Jl. Prof. Dr. Hamka Air Tawar Padang 25131, Indonesia*
Corresponding author. Email:*nurjannahbatubara73@gmail.com*

ABSTRACT

This study aims to determine the quality of critical thinking ability test instruments on the material Phase F Thermodynamics of SMA/MA in terms of validity, reliability, level of difficulty, and differentiating power. The research method used is a descriptive technique. The data consists of 31 sheets of student answers in grade XI of the 2025/2026 school year. The data were analyzed using several formulas of validity, reliability, difficulty, and differentiation. The method used to collect data is the test method. The results showed that 20 questions were valid and 1 question was invalid. The level of reliability is good where the value of the reliability coefficient r_11> r_{tabel} is 0.842 so that the question instrument is declared reliable. Meanwhile, the difficulty level of the 21 questions was stated to be moderate, the difficulty index was in the range of 0.30 < kindergarten ≤ 0.70. The difference in the good category is 2 questions, in the enough category there are about 12 questions, which is said to be bad there are about 7 questions.

Keywords: Critical Thinking Ability, validity, reliability, difficulty, and differentiation

I. INTRODUCTION

First, the user's message seems like a statement or a prompt, but it's cut off. It lists abilities for students in the 21st century: (1) think critically and creatively, (2) communicate effectively, (3) innovate, (4) find solutions to a problem, and (4) collaborate. There's a duplicate (4), probably a typo [1]. Therefore, students are required to have the ability to think critically at a deeper stage (higher order thinking) [2].

Critical thinking skills are a must in the era of digitalization and the 21st century [3]. Critical thinking skills are reflective and reasoned thinking that focuses on determining what to believe or what to do [4] . Critical thinking skills are a very important competency for students, given the challenges facing the 21st century and ongoing global changes [5] . In the learning process, critical thinking skills are very important, because students need to analyze complex concepts to overcome problems and understand natural phenomena, one of which is in physics learning.

One of the problems that is still a challenge for physics teachers in high school is the low critical thinking ability of students. This critical thinking ability is measured using an instrument in the form of an essay test. One of the instruments commonly utilized to evaluate critical thinking skills is a test. According to [6] ests serve as tools for measuring and evaluating learning in education by providing tasks or sets of tasks that students must complete, allowing educators to determine students' achievement levels he obtained score is then compared with the scores of other test takers or with a predetermined standard. Therefore, to determine the quality of the test items used, an analysis of the question items is required.

Question item analysis is an activity that must be carried out by teachers to improve the quality of the questions that have been written. The purpose of question item analysis is to determine the quality of each question item in a test in order to obtain a valid, reliable, proportional level of difficulty, and good differentiating power [7]. Quality questions are questions that can provide information as accurately as possible, so that students can know who have mastered the material and those who have not. [8] It is also argued that the purpose of

question item analysis is to find out the extent to which each question item in a test can function properly in measuring students' abilities.

A good test has characteristics and characteristics that are requirements that must be met, namely the test must be valid or have a valid level of validity [9]. A test is said to be valid if the test can accurately and correctly measure what is to be measured. Validity here, can be in the form of content validity, predictive or predictive and construction validity. In education, both test and non-test, both are instruments or tools for collecting and processing data about the variables being studied [10]. [11] explains that *Reliability* is the extent to which the results of a measurement can be trusted. An instrument is said to be reliable if the measurement results consistent and stable when used repeatedly under the same conditions. A good instrument as a good evaluation tool can be seen from several aspects, including: (1) validity, (2) reliability, (3) level of difficulty, (4) differentiating power.

Validity is related to the problem of whether the intended test question can accurately measure something to be measured. [12] says that validity can be interpreted as the extent to which the test measures what it is supposed to measure. Determining the validity of a measuring instrument can be seen as building an evidence-based argument about how well a measuring instrument measures what it should be measured [13].

Meanwhile, the reliability of the question item is related to the problem of trust. Reliability describes that a test measures something consistently that is reliable or trustworthy [14]. The relationship between validity and reliability can be explained as validity is important, while realism is necessary, because reliability supports validity [15].

With respect to the difficulty level of the question item, [16] define it as the proportion of test takers who answered the question correctly. The level of difficulty of the question item is viewed from the student's ability or ability to answer it, not from the assumption of the teacher who compiles the question, because the question item that is difficult or easy for the teacher is not necessarily difficult or easy for the student. A question item can distinguish between students who are able (to master the material asked) and students who are less capable (not yet mastering the material asked). The ability of such a question item is called differentiation (discrimination). [17] Defines that the differentiation of a question is the ability of the question item to distinguish students who score high and score low. In relation to differentiating power, a good question is a question that is answered correctly by a test taker who is able/clever/masters of the test material, and cannot be answered correctly by a test taker who has not mastered the test material [11].

Based on the description above, it is necessary to analyze the question items to determine the quality of the questions. The quality of the critical thinking ability test questions can be seen from the results of the validity, reliability, level of difficulty and differentiation tests. The researcher is interested in conducting a research with the title "Analysis of Validity, Reliability, Level of Difficulty and Differentiation of Questions on Students' Critical Thinking Ability in Thermodynamics Material Phase F SMA/MA".

II. METHOD

This study uses a type of descriptive research that aims to describe validity, reliability, difficulty, and differentiation. The research data is in the form of question sheet data and answer sheets for grade XI students of SMAN 3 Bukittinggi for the 2025/2026 school year as many as 31 people. The form of the test analyzed is an essay. Validity is a measure that indicates The level of accuracy of an instrument in carrying out its measurement function, i.e. the extent to which the measuring tool is able to accurately reveal data from the variables being studied [18]. The validity test of the question item is calculated using the formula Product Moment To find out which items are valid and which are invalid, it is necessary to correlate each item score with the total score with the correlation formula Product Moment as follows:

$$r_{xy=}\frac{n\sum \mathrm{XY}-(\sum \mathrm{X})(\sum \mathrm{Y})}{\sqrt{\mathrm{N}\sum \mathrm{X2}-(\sum \mathrm{X})2(\mathrm{N}\sum Y^{2})-(\sum Y^{2})}}$$

(Source: Ref[19]**)**

rxy = Correlation coefficient between variables X and Y
X = Total score of each question item
Y = Total score of each question
N = Number of respondents
The Validity Test Provisions, can be seen in Table 1.

Table 1. Validity Test Provisions

rxy	Criterion
rxy > 0.05	Valid
rxy < 0.05	Invalid

A valid instrument means that the measuring instrument used to obtain the data (measuring) is valid [20]**.** The valid instrument can be used to measure what should be measured. To find out whether a test is valid or not, it can be analyzed by the validity of the content *(content validity).*

A test is said to be reliable, that is, if the test can give fixed results. Therefore, the definition of test reliability is related to the problem of determining test results (Arikunto 2012). To determine the reliability of the test, a description form is used that uses the Alpha formula, namely:

$$r_{11} = [\frac{n}{(n-1)}][1- \frac{\sum {s_i}^2}{{s_t}^2}]$$

(Source: Ref[19]**)**

Information:

r11: Instrument reliability
k : The number of question items
s^2_i : Total variance
s^2 : The sum of all variances for each question

The criterion is to compare the value of r11 to the price table of the criterion price r of the current product. Where r11 is said to be significant if r11 > rtable (5% significant). The reliability coefficient criteria are found in Table 2.

Table 2. Cronbach Alpha Coefficient Criteria

No.	Reliability Coefficient	Interpretation
1.	0.80 – 1.00	Very high
2.	0.61 – 0.80	Tall
3.	0.41 – 0.60	Enough
4.	0.21 – 0.40	Low
5.	0.00 – 0..20	Very low

(Source: Ref [21]**)**

The good or bad of a test can be determined by the level of difficulty of the question. The level of difficulty of the question can be a determinant of whether a question is good or not so it needs to be revised. The difficulty index of the question item can be formulated with the following formula:

$$TK = \frac{S_A - S_B}{I_A - I_B} \qquad [19]$$

TK = Difficulty Level
S_A = Total Score of the Upper Group
S_B = Total Score of the Lower Group
I_A = Ideal Score of the Upper Group
I_B = Ideal Score of the Lower Group

Extent difficulty question get determined with use interpretation presented in Table 3.

Table 3. Difficulty Level Criteria

Difficulty Level	Category
0.00 < TK ≤ 0.30	Difficult
0.31< TK ≤ 0.70	Keep
0.71< TK ≤ 1.00	Easy

Problems that are too easy will not encourage students to try harder to solve existing problems. On the other hand, problems that are too difficult tend to make students feel hopeless because they are not able to solve them. Therefore, the ideal question is one that is classified as a medium difficulty level, so that it can

trigger optimal student motivation and involvement. A good question is a question that is in the medium classification. The difficulty index is 0.31-0.70 so, for difficult and easy questions, it should be replaced or revised. The formula used for each test item is:

$$DP = \frac{SA-SB}{IA}$$

(Source: Ref [19])

Description :

DP = Index of the differentiating power of the question item
SA = Average answer score of students in the upper
SB = Average answer score of students in the lower group
IA = Ideal maximum score
The power index of the difference in questions can be classified as shown in Table 4.

Differentiation Index	Category
DP ≤ 0.00	Very ugly
DP ≤ 0.20	Ugly
0.20< DP ≤0.40	Enough
0.40< DP ≤ 0.70	Good
0.70< DP ≤ 1.00	Excellent

(Source: Ref [19])

III. RESULTS AND DISCUSSION

Results of the Validity Test of the Question

Before the research is carried out, first make questions and answers to the questions to be validated by the supervisor. The validated questions amounted to 21 questions with valid categories. The validity of the test results can be seen in Table 5 below.

Table 5. Validation Calculation Results

Test the Validity of the Question				
No. Question	**rtable**	**Calculation**	**Conclusion**	**Information**
1	0,344	0,474	Valid	Used
2	0,344	0,4842	Valid	Used
3	0,344	0,5264	Valid	Used
4	0,344	0,4649	Valid	Used
5	0,344	0,492	Valid	Used
6	0,344	0,3564	Valid	Used
7	0,344	0,4038	Valid	Used
8	0,344	0,3422	Invalid	Revised
9	0,344	0,3511	Valid	Used
10	0,344	1,3645	Valid	Used
11	0,344	0,7617	Valid	Used
12	0,344	0,5496	Valid	Used
13	0,344	0,3647	Valid	Used
14	0,344	0,3725	Valid	Used
15	0,344	0,4956	Valid	Used

No. Question	rtable	Calculation	Conclusion	Information
16	0,344	0,5805	Valid	Used
17	0,344	0,8414	Valid	Used
18	0,344	0,5901	Valid	Used
19	0,344	0,7192	Valid	Used
20	0,344	0,6823	Valid	Used
21	0,344	0,9127	Valid	Used

Based on Table 5, it can be concluded that there are 20 valid questions out of 21 questions that have been made and tested in other schools. However, it should be admitted that the existence of 1 question that is declared invalid gives an indication of the potential for improvement that can be made. An in-depth analysis of invalid question items is important to evaluate the extent to which they meet the validity criteria and identify factors that may cause invalidity.

Results of the Reliability Test Question Item

The researcher underwent a test on critical thinking skills questions consisting of 21 Physics subject essay questions in class, XI with a total of 31 students. To determine the reliability of the research instrument, the researcher performed the calculation using the Alpha Cronbach formula. The research instrument is classified as reliable if the value of the reliability calculation exceeds or equals 0.65. Here are the results of the reliability test:

Table 6. Instrument Reliability Test Results Question

Reliability Statistics	
Cronbach's Alpha	Number of Question Items
0,842	21

Good reliability is one that shows a small error in the measurement and tested several times the results are relatively the same. If the reliability value obtained is 0.81 – 1.00 and 0.61 – 0.80 then the question is considered reliable and will be used in the research instrument, if the reliability value is sufficient, which is 0.41 – 0.60 then the question will still be used but will be revised first, and if the reliability value is low, which is 0.21 – 0.40 and 0.00 – 0.20 then the question will not be used.

Difficulty Test Results of Question Items

The researcher underwent a test on critical thinking skills questions consisting of 21 Physics subject essay questions in class, XI with a total of 31 students.

Table 7. Results of Calculation of the Difficulty Level of the Trial Question

Question No.	Rate Coefficients Difficulty	Category
1	0,6935	Keep
2	0,6854	Keep
3	0,7016	Keep
4	0,6935	Keep
5	0,7016	Keep
6	0,7016	Keep
7	0,6693	Keep
8	0,6693	Keep
9	0,6290	Keep

Question No.	Rate Coefficients Difficulty	Category
10	0,5645	Keep
11	0,6612	Keep
12	0,6854	Keep
13	0,6935	Keep
14	0,6451	Keep
15	0,5967	Keep
16	0,5483	Keep
17	0,6129	Keep
18	0,5725	Keep
19	0,6290	Keep
20	0,6451	Keep
21	0,6532	Keep

Based on table 9, of the 21 question items that have been tested, 20 question items were obtained that were declared valid. The value of the reliability coefficient obtained was 0.842 included in the very high reliability category, the difficulty level of the 21 question items that were tested was obtained all question items including medium criteria, there were 12 questions with sufficient category differences and 7 questions in the bad category, then 7 questions in the bad category were revised. It can be concluded that the questions are worthy of being tested in this study.

Discussion

This study aims to analyze essay questions designed to measure students' critical thinking skills**,** with assessment guidelines compiled based on the stages of critical thinking in Class XI Physics at SMA Negeri 3 Bukittinggi for the 2025/2026 Academic Year. The quality of the question items is seen from the characteristics of the assessment of the question items which consist of: validity, reliability, level of difficulty, differentiation of the question. The following is a discussion of the characteristics of each question item assessment:

ACKNOWLEDGMENT

Please write a thank you note in this section. Gratitude is addressed to research funders or parties who contributed to the implementation of research or writing articles, other than the author.

1. Validity

   The validity of the question items in this study is seen from the calculation and then consulted with the product moment table at a significant level of 5% according to the number of test participants. The number of students who took part in the trial of Physics class XI at SMA Negeri 3 Bukittinggi was 31 students. Thus N=31 which shows a value of 0.344. The criteria used in the interpretation of the validity of the question item is > 0.344 means valid and if < 0.344 then the question is invalid.$r_{tabel} r_{tabel} r_{tabel}$

   According to [22] that one of the characteristics of a good critical thinking ability test is to have validity. A critical thinking ability test with high validity can be said to be reliable and there is no need to doubt its accuracy in measuring students' critical thinking skills. [23] It also argues that a good test must have the characteristics of validity in order to present accurate information about the condition of the students who take the test. This information is very useful for handling the students concerned. If a test is invalid, the information obtained by the teacher based on the results of the learning test will be misleading and detrimental to both the teacher and the student.

   The results of the analysis of the Class XI Critical Thinking Ability Questions at SMAN 5 Bukittinggi with 21 essay questions that were declared valid amounted to 20 items while 1 item was invalid. Research conducted [24] with the title "Analysis of Critical Thinking Skills of High School Students on Static Fluid Materials" with the results of the analysis: valid question items 9 questions, and invalid 6 questions. In comparison, the question of the ability to think critically about thermodynamic materials is better than the question of static fluids from the validity of the question.

From the description above, it can be concluded that the Class XI Thermodynamics Material Critical Thinking Ability question is a good quality question in terms of validity because the number of valid questions is 20 items. This is in accordance with the theory that question items that have high validity have reliability in measuring student learning outcomes.

The criteria for question items are good based on validity, if the question is said to be valid. So a good question item is 20 items with question numbers: 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21.

Follow-up to the results of the analysis of question items as follows:

a. Invalid question items are declared as dropped questions and should be discarded, but if they are to be reused, they should be revised.
b. Valid question items can be reused and included in the question bank.

2. Reliability

Question reliability is the level of consistency in measuring learning outcomes [25]**.** The critical thinking ability test is said to be consistent if it provides fixed measurement results when tested many times on the same group at different times. Reliability is seen from the Alpha value on the output after the calculation is made.

The results of the analysis showed that the Critical Thinking Ability Questions at SMAN 3 Bukittinggi had a reliability index of 0.842 (Alpha value). Based on the criteria used, the reliability index is included in the very high category. Research conducted [24] with the title "Analysis of Critical Thinking Ability of High School Students on Static Fluid Materials" obtained the results of a reliability analysis of 0.82. When compared, both questions have very high reliability.

So it can be concluded that the Critical Thinking Ability at is of good quality judging from the reliability of the questions. This is in accordance with the theory [26] that "A reliable test is if it has a high coefficient and a standard measurement error (standard error of measurementnment) low". One of the characteristics of questions has high reliability if the test consists of many question items with valid categories. In addition, the high and low reliability index is influenced by several factors, namely test length, score distribution, difficulty level, and objectivity [26]**.**

3. Difficulty Level

Good question items have a moderate level of difficulty in the sense that they are not too difficult and not too easy. Problems that are too easy do not stimulate students to solve problems. On the other hand, questions that are too difficult will cause students to not have enthusiasm in doing the questions because they are beyond the reach of the student's ability.

The results of the difficulty level analysis of the Class XI Thermodynamics Material Critical Thinking Ability Questions show that all questions are suitable for use in the medium category. This is in line with the research conducted [27] that Questions with a moderate level of difficulty are ideal questions and are suitable for use because they are able to distinguish students' abilities optimally.". The difficulty level index of good question items is between 0.31-0.70. So it can be concluded that the Critical Thinking Ability Questions from the level of difficulty are of quality because 21 questions are used.

The criteria for concluding the quality of the question items are good based on the level of difficulty, including moderate questions. Good question items amounted to 21 items, namely numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21.

According to [22] The follow-up that can be done after the question item is analyzed for its difficulty level is as follows:

a. Items that, based on the analysis results, fall into the good category (moderate difficulty level) can be directly entered into the question bank.
b. Items in the category are too difficult, there are three possible follow-ups that can be done, namely: (1) items are discarded or dropped and not removed again in the next learning outcome test; (2) re-researched, tracked and traced to the cause of items that are difficult to answer by the testee. After that, improvements are made so that item items can be reused in learning outcome tests; (3) It is used in tests that are very strict (selection tests) so that they can be stored in a separate question bank.
c. Item items in the category are easy, there are also three possible follow-ups, namely: (1) item items are discarded or dropped and are no longer removed in the learning outcome test; (2) re-examined, tracked and traced to find out the causative factors of the item items can be answered correctly by almost all testees. After it is known to be fixed, the item in question is tried to be removed again to find out whether the item's difficulty level is better or not; (3) used in tests that are loose, in the sense

that most of the testees are declared to have passed the selection test. In this condition, it is very wise if items in the category are easily excluded in the selection test.

4. Differentiating power

Differentiating power is the ability of question items to distinguish between smart students and less smart students. The results of the analysis showed that of the 21 items used by the researcher, the questions had a category of 12 enough items, both 2 items, and 7 bad items. Research conducted by [24] with the title "Analysis of Critical Thinking Skills of High School Students in Static Fluid Materials" with the results of the analysis of the questions having very poor power of 1 item, 9 items in the bad category, and 5 in the medium category. When compared to the ability to think critically, thermodynamic materials, the differentiating power of the question is better than the question of static fluids.

According to [22] Knowing the differentiating power of question items is important, because one of the bases in compiling learning outcome test questions is the assumption that the ability of one testee is different from one another and that the learning outcome test question item must be able to provide test results that illustrate the difference in ability among testees. A good question is a question item that can distinguish between students who are good and students who are not good at it, in this case the question can be answered correctly by good students.

From the description above, it can be concluded that the Questions on Critical Thinking Ability in Thermodynamics Materials Class XI SMAN 5 Bukittinggi for the 2025/2026 Academic Year are seen in terms of the differentiating power of questions, including good quality. Of the 21 questions that were considered bad, only 7 were bad. In the sense that there are many questions that can distinguish between students who master the material and students who do not master the material.

The criteria for the quality of good question items seen from the differentiating power of the question are questions categorized as very good, good, and sufficient. The number of question items that are of high quality is enough to have 12 items with item numbers 1, 3, 4, 5, 7, 13, 14, 15, 16, 19, 20, 21.

According to [22] The follow-up of the question items after analyzing the differentiating power is as follows:

a. Items that have a good differentiating power are stored in the question bank. These items can be reissued during future learning outcome tests.
b. Items with low differentiating power, there are two possibilities that will not continue, namely: (1) be traced to be improved and then reused in future learning outcome tests to find out whether the differentiating power has increased or not. (2) Dumped (dropped).
c. Items with a negative discrimination index number should be discarded because the quality of the question items is very poor.

CONCLUSION

Based on the results of the analysis of students' critical thinking ability instruments in the Thermodynamics Phase F SMA/MA material reviewed from the aspects of validity, reliability, level of difficulty, and differentiation, it can be concluded that in general the test instruments developed have good quality and are suitable for use as an evaluation tool for students' critical thinking skills. The reliability value of the test is included in the strong category**,** so it can be trusted to measure critical thinking skills stably. In terms of difficulty, the question items are spread out in medium, showing that this test is able to distinguish students' abilities well. Meanwhile, the results of the differential analysis showed that most of the question items had a good ability to distinguish between students who mastered the concept and those who did not understand the material in depth.

ACKNOWLEDGEMENT

Please write a thank you note in this section. Gratitude is addressed to research funders or parties who contributed to the implementation of research or writing articles, other than the author.

REFERENCES

[1] Whitby, G. B. (2007). Introduction: Why new pedagogies? Strands of relevance. ACEL 2007 International Conference Sydney, Australia, 1–11.

[2] Istiyono, E. (2014). Developing Higher Order Thinking Skill Test Of Physics (Physthots) For Senior High School Students. Educational Research and Evaluation, 18(1), 1–12.

[3] Mahardini, T., Khaerunisa, F., Wijayanti, I. W., & Salimi, M. (2019). Research Based Learning (Rbl) To Improve Critical Thinking Skills. Social, Humanities, and Educational Studies (SHES): Conference Series, 1(2), 466. https://doi.org/10.20961/shes.v1i2.26816

[4] Ennis, R. H. (2011*).* The Nature of Critical Thinking. Informal Logic, 6(2), 1–8. https://doi.org/10.22329/il.v6i2.2729

[5] Indahwati, S. D., Rachmadiarti, F., & Hariyono, E. (2023). Integration of PJBL, STEAM, and Learning Tool Development in Improving Students' Critical Thinking Skills. IJORER: International Journal of Recent Educational Research, 4(6), 808–818. https://doi.org/10.46245/ijorer.v4i6.434

[6] Sudijono. (2012). Introduction to Educational Evaluation. Jakarta: Pt. Raja Grafindo Persada.

[7] Arikunto, S. (2013). Basics of Educational Evaluation**.** Jakarta: Bumi Aksara.

[8] Fitriani, R., & Anshori, I. (2019). Analysis of Multiple-Choice Questions for the Final Semester Exam of Fiqih Subjects at MTsN 1 Lamongan Academic Year 2018/2019. *Journal of Education and Learning*, 8(1), 45–54.

[9] Naga, D. S. (2013). Score Theory on Education Measurement**.** Jakarta: Gunadarma.

[10] Mardapi, D. (2017). Measurement, Assessment, and Evaluation of Education. Yogyakarta: Parama *Publishing.*

[11] Azwar, S. (2012). Reliability and Validity**.** Yogyakarta: Student Library.

[12] Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research**.** International Journal of Academic Research in Management (IJARM), 5(3), 28–36.

[13] Sullivan, G. M. (2011). A primer on the validity of assessment instruments**.** Journal of Graduate Medical Education, 3(2), 119–120. https://doi.org/10.4300/JGME-D-11-00075.1

[14] Bajpai, S., & Bajpai, R. (2014). Goodness of measurement: reliability and validity. International Journal of Medical Science and Public Health, 3(2), 112-116.

[15] Sukiman. (2012). Development of Learning Evaluation System. Yogyakarta: And then there is Madani.

[16] Boopathiraj, C., & Chellamani, K. (2013). Analysis of test items on difficulty level anddiscrimination index in the test for research in education. International journal of social science & interdisciplinary research, 2(2), 189-193.

[17] Kocdar, S., Karadag, N., & Sahin, M. D. (2016). Analysis of the Difficulty and DiscriminationIndices of multiple-choice questions according to cognitive levels in an open and

[18] Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric Theor**y** (3rd ed.). New York: McGraw-Hill.

[19] Sundayana, R. (2020). Educational Research Statistics: Alphabet.

[20] Sugiyono. (2013). Quantitative, Qualitative, and R&D Research Methods (Issue January).

[21] Guilford, J. P. (2005). Fundamental Statistics in Psychology and Education (4th ed.). New York: McGraw-Hill Book Company.

[22] Anas Sudijono. (2011). Introduction to Educational Evaluation. Jakarta: Raja Granfindo Persada.

[23] Kuseiri and Suprananto. (2012). Measurement and Valuation Education.Yogyakarta: Graha Ilmu.

[24] Novianty, Amalia. 2022. Analysis of Critical Thinking Skills of High School Students on Static Fluid Materials

[25] Sumarna Supranata. (2005). Analysis, Validity, Reliability and Interpretation of Results

[26] 2004 Curriculum Implementation Test. Bandung: Remaja Rosdakarya.

[27] Zainal Arifin. (2013). Learning Evaluation. Bandung: Remaja Rosdakarya.Supriatna, D. (2018). Analysis of Multiple Choice Questions for Science Subject Class VIII. Journal of Education and Evaluation, 6(2), 45–53