e-ISSN: 2721-1274
p-ISSN:2721-2718

PORTRAIT OF STUDENTS' CRITICAL THINKING ON WORK AND ENERGY MATERIALS

Aminah, Wawan Bunawan
Physics Education, Universitas Negeri Medan, Indonesia
aminah.natal27@gmail.com

ABSTRACT

This research aims to develop a critical thinking test instrument on the material of work and energy that completes the aspects of validity, reliability, level of difficulty, item discrimination, and distractor effectiveness and maps students' critical thinking skills. The type of research used is quantitative research with the Research and Development method using the Tjeerd Plomp educational design research model integrated with the ADDIE development model with 5 stages, namely analysis, design, development, implementation, and evaluation. The instrument developed consists of 20 multiple-choice questions. The test sheet consists of multiple-choice physics test items arranged based on critical thinking indicators according to Facione. The results of the content validation test, according to material, construction, and language experts, obtained an average value of 0.97 with a high validity category. The validation results using the Cohen Kappa formula obtained a Kappa value of 0.63 with a "strong" category. Data analysis using the Microsoft Excel application. The results of the small class test obtained 75% valid items, high reliability with a value of 0.82, 50% moderate difficulty level, 60% of the category item discrimination was sufficient with an average of 0.38, and 85% of the effectiveness of the good category distractor. The results of the large class test obtained 80% valid items, medium reliability with a value of 0.62, 70% of the difficulty level of the medium category, the average item discrimination of 0.32 categories was sufficient, and 95% of items with good category distractor effectiveness.. The final results of the development of Critical Thinking test instruments on Business and Energy materials at PAB 8 Saentis Private High School using the ADDIE model R&D method amounted to 16 test items with the categories of valid, moderate reliability, good differentiation, moderate difficulty level, and good distractor effectiveness. Thus, the test instruments developed in the study are said to be feasible and function to map students' critical thinking skills

Keywords: Critical Thinking; Facione; Multiple Choice; Work; Energy

INTRODUCTION

Critical thinking skills are needed to understand a problem so that information can be understood correctly (Munawwarah et al., 2020). The government has made several efforts to improve critical thinking skills, including teacher training in higher-level learning, and training in the preparation of higher-level thinking questions (Sani, 2019). Teachers' skills in compiling questions are

e-ISSN: 2721-1274
p-ISSN:2721-2718

indispensable for improving the learning process. The question is a HOTS-oriented assessment instrument to determine students' high-level thinking skills. The assessment instrument will answer the desires of the K-13 curriculum because as many as 50% of the Basic Competencies require high-level thinking skills (C4, C5, and C6) (Desilva et al., 2020: 42). Among the high-level thinking is critical thinking.

The development of critical thinking test instruments must meet the feasibility of good test instruments. A test is a good measurement tool if it meets the requirements, namely, validity, reliability, objectivity, practicality, and economy. A test is valid if it can accurately measure what is being measured. The test is said to be reliable if it gives a fixed result if it is tested many times (Arikunto, 2013). Based on the researcher's findings, the development of this test instrument is needed at SMAS PAB 8 Saentis.

SMAS PAB 8 Saentis is one of the private high schools located at Jalan Kali Serayu PTP II Saentis, Saentis Village, Percut Sei Tuan District, Deli Serdang Regency, North Sumatra Province. Based on the results of pre-research interviews at SMAS PAB 8 Saentis, teachers used a problem-solving learning model. In compiling the questions, teachers have not used Critical Thinking-based questions, the instruments used are still higher-order thinking (HOT) cognitive problems that tend to involve the ability to remember, understand, and apply. The activeness of students in learning is also uneven because they are not interested in lessons, especially Physics. From the explanation above, it can be seen that it is necessary to develop a Critical Thinking test instrument. In improving critical thinking skills, the development of physics test instruments that can train critical thinking skills that meet the feasibility of good test instruments can be used, therefore researchers research the development of critical thinking-based test instruments. The materials used in the research are business and energy materials. The purpose of this study is to produce critical thinking-based test instruments, to find out the characteristics of critical thinking-based test instruments, and to map students' critical thinking skills on work and energy materials.

METHODS

This type of research is developed with a model of R&D (Research and Development) development research approach. This research was carried out at SMA PAB 8 Saentis Jalan Kali Serayu Dusun 16, Percut Sei Tuan District, Deli Serdang Regency, North Sumatra Province, and was carried out in February-March 2023. The subjects in this study are grade XI students at SMA PAB 8 Saentis for the 2021/2022 Academic Year consisting of class XI MIA 1 as a research sample.

The development model used in this study is the ADDIE Learning design model is integrated with the steps of Educational Design Research recommended by Tjeerd Plomp based on the consideration that the educational design research model is relevant for educational practice because it aims to develop research-based solutions to complex problems or to develop or validate theories about the learning process as well as develop instructional/learning model products that are right on target, effective and dynamic and very helpful in the development of teachers (Plump, 2013). The scheme of these two models is arranged sequentially

e-ISSN: 2721-1274
p-ISSN:2721-2718

and programmatically so that the sequence of activities is systematic shown in Figure-1.

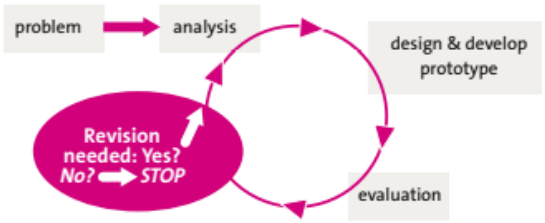

Figure-1. Design of Research Model

Whatever the purpose of design research, the research process always incorporates a systematic educational design process, such as the scheme above. Therefore, all systematic educational and instructional design processes are cyclical: analysis, design, evaluation, and revision activities are repeated until the right balance between the intended ideals and the realization has been achieved (Sugiyono, 2017). The systematics of educational design research with the ADDIE development model can be seen in Figure-2.

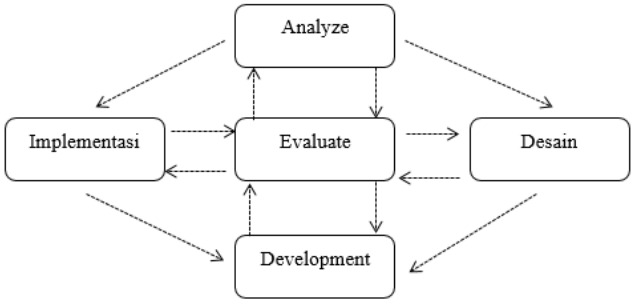

Figure-2. ADDIE Models for Product Development

The data collection technique is to conduct interviews with teachers to find out the state of the school and students where the research is conducted and test the feasibility of the instrument using assessments from experts who will prove the validity of the content of the instrument. Conducting tests of the developed instruments to students and conducting data analysis. All data analysis uses Microsoft Excel.

Data analysis is related to validity, reliability, level of difficulty, differentiation, and effectiveness of the distractor. The validity of the question items is calculated by the product moment correlation formula by Karl Pearson.

$$r_{xy} = N = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{\{N\sum x^2 - \sum x^2\}\{N\sum y^2 - \sum y^2\}}} \quad (1)$$

To determine the Reliability value, the Kuder and Richardson Engineering formula (K-R 20) is used:

$$r_{11} = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum \sigma_i^2}{\sigma_1^2}\right) \quad (2)$$

The criteria for a reliable question item are the following values:

e-ISSN: 2721-1274
p-ISSN:2721-2718

Table-1. Reliability category range

Reliability Range	Category
1.00	Perfect
0.90 – 1.00	Very High
0.70 – 0.90	High
0.40 – 0.70	Average
0.20 – 0.40	Low

To measure the difficulty level of each multiple-choice question item, the following formula is used:

$$I = \frac{B}{J} \quad (3)$$

The criteria for the difficulty index of the questions can be seen in the following table:

Table-2. Difficulty index criteria

Difficulty Index Range	Criteria
P≤ 0,3	Difficult
$0,3 < P \leq 0,7$	Average
$P \geq 0,7$	Easy

To determine the differentiation of a test, the following formula is used:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B \quad (4)$$

The criteria for the differentiation of each question according to Daryanto (1999) are as follows:

Table-3. Criteria of differentiation questions

Differentiation Range	Category
0.00 – 0.20	Poor
0.21 – 0.40	Satisfactory
0.41 – 0.70	Good
0.71 – 1.00	Excellent
If D is negative	Not Good

The analysis of the effectiveness of the distractor is used to determine whether the distractor has functioned properly or not. A cheater is said to function well if chosen by 5% of test takers (Arikunto, 2013).

RESULT & DISCUSSION

The results of this research are in the form of a Critical Thinking Test instrument in the form of multiple-choice questions on Business and Energy material in Senior High School (SMA). This instrument consists of 20 questions taken from several sources and then developed by researchers. The preparation of the product is carried out through the stage of basic competency analysis,

e-ISSN: 2721-1274
p-ISSN:2721-2718

preparation of grids, preparation of questions, testing of questions, revision of questions, and determination of answer keys and score scores.

Furthermore, critical thinking test instruments are developed with the following characteristics: validation, reliability, differentiation, and difficulty levels that are tested qualitatively and quantitatively. The qualitative test was carried out through an expert validation test consisting of material, construction, and language aspects. Quantitative tests were carried out on small-class and large-class tests by the Plump educational design research method and the ADDIE development model.

The content validation test on the test instrument was carried out by three experts with an average validity value of 0.97 with the validity statement "High" as stated in Table-1. Validation tests were also carried out by assessing validator agreement using the Cohen Kappa formula. The results of the test using Cohen Kappa obtained a Kappa value of 0.63 which can be categorized as having a "Strong" agreement. Validation quantitative data was processed using Microsoft Excel to determine the equation of the validation index of question items (Index V) proposed by Aiken. The results of the validation test are shown in Table-4.

Table-4. Validation of data by expert

No	Validation Aspect			Item Personal Validity	Information
	Material	Construct	Language		
1	1	1	1	1	High
2	0,98	1	1	0,99	High
3	0,96	0,95	1	0,96	High
4	0,94	0,95	1	0,95	High
5	1	0,97	0,95	0,96	High
6	1	0,98	0,98	0,98	High
7	0,98	0,98	0,98	0,97	High
8	0,98	0,93	0,98	0,95	High
9	0,98	0,95	1	0,97	High
10	0,98	0,98	0,98	0,97	High
11	0,98	0,95	1	0,97	High
12	0,98	0,95	0,96	0,95	High
13	0,96	1	1	0,98	High
14	0,98	0,98	0,98	0,97	High
15	0,98	0,95	1	0,97	High
16	1	1	1	1	High
17	1	1	1	1	High
18	0,98	0,98	0,96	0,96	High
19	0,96	0,97	1	0,97	High
20	0,96	0,97	1	0,97	High
Average				0,97	High

In the information in the expert validation questionnaire, several instruments need to be revised. So, the test instrument is then revised according to expert advice and then tested to the small class test stage.

e-ISSN: 2721-1274
p-ISSN:2721-2718

A. Small Class Test

The small group trial was carried out at PAB 8 Saentis Private High School class XI with 15 students within 120 minutes. In this test, it is expected that the test instrument meets all aspects of testing, namely validation, reliability, level of difficulty, and differentiation. The validity of each question item is calculated by creating a score table. Questions that are answered correctly are given the number 1, and questions that are answered incorrectly are given the number 0. While the total score is the total number of question item scores. The validation value was obtained by using the product-moment correlation of the rough number using the Microsoft Excel application. The results of the small class validation test obtained 15 valid instruments.

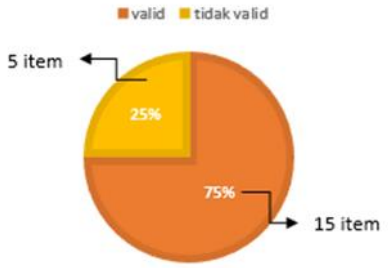

Figure-3. Validity of small class test

The reliability of the Critical Thinking test instrument was analyzed using the Richadson Kuder formula (KR-20) with the help of the Microsoft Excel application. The reliability of the test instrument in the small class test was categorized as medium with a value of 0.82.

The results of the analysis of the difficulty level in the 20 items tested were 10 questions with the easy category, 10 questions with the medium category, and no difficult items as seen in Figure-4.

Figure-4. Difficulty item spread

The results of the item discrimination analysis obtained 13 test items with good categories, 3 items with sufficient categories, 2 items categorized as bad, and 2 items categorized as bad. Test items with good and sufficient categories are acceptable without revision so that they can be used for the next stage of the exam. Test items with bad and bad categories need to be revised. The percentage of differential power test results can be seen in Figure-5.

e-ISSN: 2721-1274
p-ISSN:2721-2718

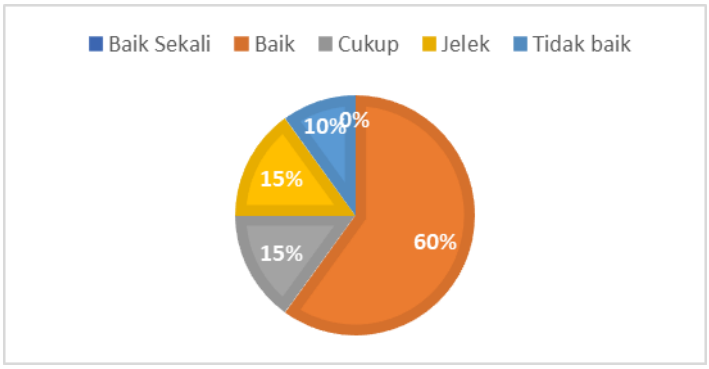

Figure-5. Differentiation for small class test

The results of the effectiveness test of the distractor from 20 test items, it was found that 17 test items (85%) had good and very good distractors so that no revision was made. Meanwhile, 2 items that have sufficient effectiveness and 1 item have poor effectiveness need to be revised so that they can be continued to the large-class test stage. The percentage effectiveness of the deception can be seen in Figure-

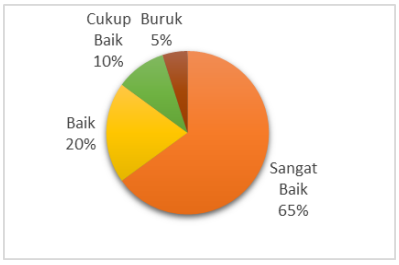

Figure-6. Distribution of the effectiveness of deception in the Small Class Test

Before proceeding to the Large Class Test, several instruments must be revised first based on the test results in the small class. The recapitulation of the results of the small class test can be seen in Table-5.

Table-5. Recapitulation of Small Class Test Results

Number of questions	Criteria that fulfilled	Item criters test	Revision
4,5,7,8,9,11,16,18,19,20	4	Very good	Nothing
1,2,14,15,17	3	Good	Minor
3,6,10,12,13	2	Enough	Mayor
	1	Not bad	Throw away
	0	Very not good	Throw away

Of the 20 test instrument items tested, 10 items were accepted without revision, 5 were accepted with minor revisions and 5 items were accepted with major revisions.

B. Large Class Test

The revised test instrument was then tested in a large class was carried out on 30 students with a training time of 120 minutes. The large class test aims to

e-ISSN: 2721-1274
p-ISSN:2721-2718

obtain information about the level of students' critical thinking skills, and the instrument is expected to meet all aspects of the test, namely validation, reliability, difficulty level, quality of item discrimination, and effectiveness of the deceitful.

The results of the validation of the test instrument in the large class test were obtained 16 valid items and 4 invalid items. The validity percentage can be seen in Figure-7.

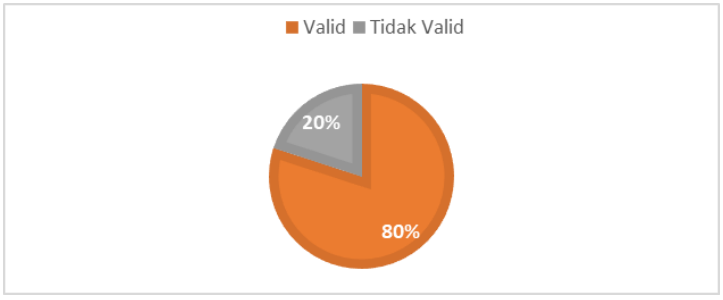

Figure-7. Validity of big class test

The results of the reliability analysis obtained a reliability value of 0.604 with the medium reliability category. The results of the difficulty level analysis were obtained for 14 medium category items and 6 difficult category items. The percentages are depicted in Figure-8.

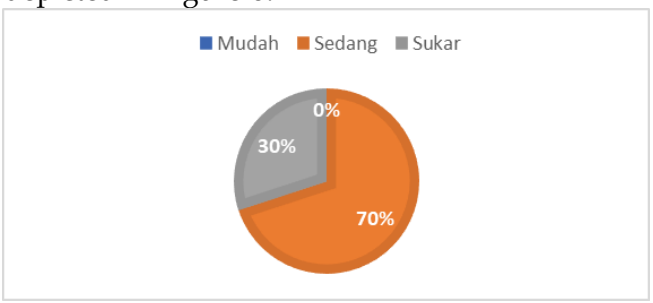

Figure-8. Reliability of big class test

The difference in the research of test instruments with a very good category of 1 item, a good category of 4 items, a sufficient category of 11 items, and a bad category of 4 items. The percentage can be seen in Figure-9.

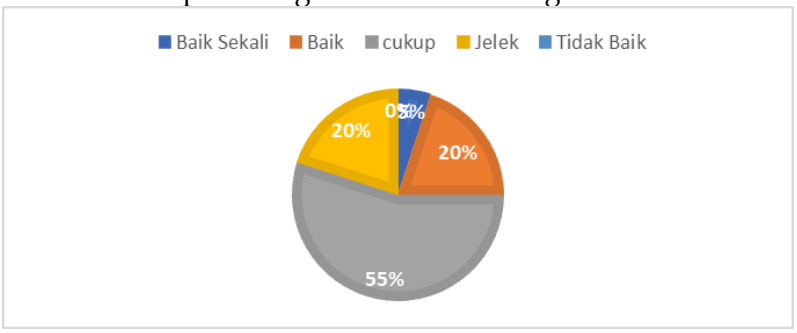

Figure-9. Difficulty item spread of big class test

The results of the analysis of the effectiveness of the distractor show that of the 20 items, 19 tests have very good and good distractors so no revision is carried out. Items that have enough effectiveness are 1 item and need to be revised. The

e-ISSN: 2721-1274
p-ISSN:2721-2718

percentage of effectiveness of the distractor was obtained by 75% of the questions were very good, 20% of the questions were good, 5% of the questions were sufficient, and 0% were bad as can be seen in Figure-10.

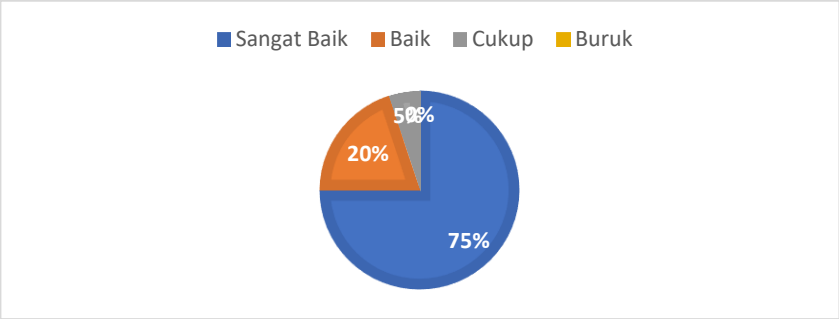

Figure-10. Effectiveness of big class test

The results of the large class test are summarized in the recapitulation of the large class test listed in Table-6.

Table-6. Recapitulation of test results in big class test

Number of Questions	Criteria Fulfilled	Criteria of Item Test	Revision Information
1,7,8,11	4	Very Good	Nothing
2,4,5,9,10,12,13,14,15,16, 17,20	3	Good	Minor
3,6,18,19	2	Enough	Mayor
	1	Not Good	Thrown Away
	0	Very Not Good	Thrown Away

In the results of the second trial, there was an increase in valid test items by as much as 5%. The increase in validation test results can be due to revised test items and the increasing number of participants. However, there was a decrease in test reliability from a value of 0.82 categorized as high to a value of 0.604 categorized as moderate. In the small class test, it was found that of the 20 instruments tested, there were 10 items classified as easy and 10 items classified as moderate with a percentage of difficulty level in the large class test namely, there were 6 items classified as difficult and 14 items classified as moderate and no items classified as easy. The item discrimination has different results were in the large class tests there is an increase for the very good category, and there is a decrease for the bad and bad categories. This shows that the test instrument is can be distinguishing test takers who have critical thinking skills from test takers who do not have critical thinking skills. The analysis of the effectiveness of the distractor also underwent changes and increases which showed that the number of test participants could affect the effectiveness of the distractor.

C. Mapping of critical thinking skills based on indicator per item

The results of mapping students' ability to think critically are made based on a grid of instruments, namely by looking at the number of indicators that students can master in the small class test and the large class test. The mapping can be seen in Figure-11.

e-ISSN: 2721-1274
p-ISSN:2721-2718

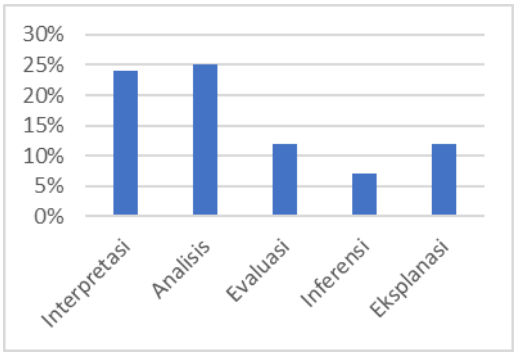

Figure-11. Mapping of student critical thinking skill based on Indicators per item

D. Mapping of Critical Thinking skills Based on Stages on Students' Answer Sheets

In addition to mapping students' skills based on critical thinking indicators in the questions, mapping was also carried out based on the critical thinking stages found on the students' answer sheets. Table 7. shows the number of stages that students were able to complete on each question item.

Table 7. Number of students at each CT stage

No Soal	Critical thinking stages				
	Interpretation	Analyze	Inference	Evaluation	Eksplanationn
1	18	14	13	0	0
2	21	17	17	2	1
3	13	10	6	0	0
4	12	9	10	4	0
5	1	0	0	0	0
6	18	17	17	3	0
7	3	1	1	0	0
8	11	11	6	1	0
9	4	0	0	0	0
10	5	4	4	2	2
11	6	6	6	2	1
12	3	2	0	0	0
13	3	2	2	0	0
14	3	2	0	0	0
15	3	2	0	0	0
16	4	3	1	0	0
17	3	2	0	0	0
18	3	2	0	0	0
19	12	12	12	6	4
20	2	1	0	0	0
Total	148	117	95	20	8
Rata-rata	7,4	5,85	4,75	1	0,4

e-ISSN: 2721-1274
p-ISSN:2721-2718

If the table above is presented in the form of a diagram, the percentage of the table can be seen in Figure 4.10 below.

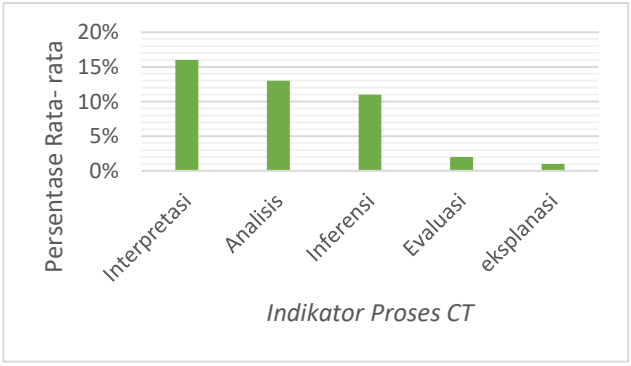

Figure 11. CT percentage diagram at critical thinking stages

Based on Figure 4.10, it is obtained that the average number of students who are capable at the interpretation stage is 16%, the analysis stage is 13%, the inference stage is 11%, the evaluation stage is 2%, and the explanation stage is 1%.

Based on Figure 4.9 and Figure 4.10, there is a difference in the percentage of students' critical thinking skills. Figure 4.9 shows a diagram that goes up and down for each critical thinking indicator contained in the question. In the mapping based on the student's answer sheet, a decreasing diagram was found. When compared, this difference explains that not all students answered the answer choices test in the question according to their critical thinking skills.

Based on the mapping results above, a scale of satisfaction with students' critical thinking skills can be presented in the form of a table using the Likert Scale with the following indicators.

Table 8. Satisfaction Scala

Persentase	Keterangan
0%-20%	Unsatisfactory
21%-40%	Less Satisfactory
41%-60%	Quite Satisfying
61%-80%	Satisfying
80%-100%	Very Satisfactory

The results of students' critical thinking mapping are in the range of 0%-20%, so that based on table 8, it can be seen that students' critical thinking skills are categorized as unsatisfactory. The final results of the research on the development of the critical thinking test instrument on the material of work and energy at SMA Swasta PAB 8 Saentis amounted to 16 test items using the ADDIE R&D model method with valid categories, moderate reliability, good discrimination power, moderate level of difficulty, and good distractor effectiveness. Thus, the test

e-ISSN: 2721-1274
p-ISSN:2721-2718

instrument developed in the study is said to be feasible and can map students' critical thinking skills.

CONCLUSION

Based on the results of the research, the development of a physics test instrument based on Critical Thinking on Business Materials and Energy at the High School Level was developed in the form of multiple choice 20 questions. The conclusion of the analysis and discussion is as follows:

1. The development of test instruments through a series of procedures, namely: Defining Critical Thinking skills, formulating the construct and objectives of developing test instruments, compiling test item formats, making scoring guidelines, validation tests by 3 experts with validation data analysis using Aiken V and Cohen Kappa, making revisions according to expert suggestions, conducting small class tests and large class tests with validation tests using product moment correlation, reliability test using the Kuder and Richadson Technique (K-R 20), testing the level of difficulty, testing the item discrimination and testing the effectiveness of the trickster.
2. The number of questions with the characteristics of the Critical Thinking test instrument developed are:
    - Valid according to material, construction, and language experts with a high category with an average score of 0.97 and supported by the Cohen Kappa test with a score of 0.63 is categorized as "strong agreement" between assessors and shown by empirical evidence through item validation with 75% (15 items) valid on the small class test and 80% (17 items) valid on the large class test.
    - Reliable with a score of 0.82 in the high category ($r \geq 0.70$) for the large class test results and 0.60 in the medium category ($r \geq 0.40$) for the large class test results.
    - The difference between the test items is 60% (12 items) in the good category in the small class test, and the difference with 25% (5 items) in the good category in the large class test.
    - The difficulty level of medium category test items is 50% (10 items) in the small class test and 70% (14 items) of the medium category test in the large class test.
    - Excellent and good test item cheater effectiveness 85% (17 items) in small class test results and 95% (19 items) test item cheaters with very good and good category test items in large class test results.
    - The developed test instrument can map students' critical thinking skills with two mappings, namely mapping based on indicators for each test item and mapping based on students' answer sheets. The mapping results obtained the range of students' critical thinking skills is 0%-20% with an unsatisfactory category.

BIBLIOGRAPHY

Arikunto. S. (2013). Dasar- dasar Evaluasi Pendidikan : Edisi Kedua. Jakarta : Bumi Aksara.

e-ISSN: 2721-1274
p-ISSN:2721-2718

Desilva, D., dkk. (2020). Pengembangan Instrumen Penilaian Hasil Belajar Fisika Berorientasi HOTS (Higher Order Thinking Skills) Pada Materi Elastisitas Dan Hukum Hooke. Jurnal Kumparan Fisika,3(1),41-50.

Plomp, T., et al. (2013). Educational Design Research. Enschede : SLO

Sani, R.A. (2019). Pembelajaran Berbasis HOTS (Higher Order Thinking Skills). Tangerang : TS Smart.

Sugiyono. (2017). Metode Penelitian Kuantitatif, Kualitatif dan R & D. Bandung : Alfabeta.