Development of Test Instruments for Problem Solving Ability of Work and Energy Materials

Mustika Wati[1], Rahmawati[2], Saiyidah Mahtari[3], Dewi Dewantara[4], Yasmine Khairunnisa[5], Theo Jhoni Hartanto[6]

{mustika_pfis@ulm.ac.id[1]}

[1,2,3,4,5]Physics Education, Lambung Mangkurat University
[6]Physics Education, University of Palangka Raya

Abstracts. This research aims to produce a problem-solving ability test instrument on work and energy materials. This research is a Research and Development (R&D) with the Borg & Gall model. The subjects of the research trial were 176 students of class X MIPA from three schools. Data analysis includes design validation, instrument validity, reliability, difficulty level, and instrument separation index. Design validation by 6 validators resulted in a valid and highly reliable test instrument. In the product try out (small group), with a total of 29 students, (1) 17 questions were valid with a very high reliability of 0.92, (2) the level of difficulty is easy, medium, and difficult criteria, (3) the separation index of very bad criteria question was discarded as many as one question. In the field test, conducted in 3 schools with as many as 176 students, the data showed that (1) out of 16 questions, 5 questions were selected as the final product, (2) the results of the validity of the instrument obtained valid category, (3) it had reliability with sufficient criteria of 0.69, (4) the level of difficulty is in medium criteria for all questions, (5) the separation index of questions with bad criteria was reused based on variable map items. So, the final results obtained by the problem-solving ability test instrument amount to 5 questions and can be used as a test of student learning outcomes.

Keywords: Test instrument, problem-solving ability, Rasch model, work and energy.

1 Introduction

The challenge in the implementation of higher education in the 21st century is to prepare human resources who have problem-solving competencies [1]. Problem-solving ability is the ability of students to solve problems through a series of activities to find information from various learning sources and connect new knowledge gained with knowledge and experience that has been previously possessed [2]. Problem-solving skills are needed in learning physics because problems in physics are complex problems and are related to everyday life [3], [4].

Several factors affect the low problem-solving ability of students. Arief [5] suggests that students are less able to understand basic concepts and tend to memorize formulas. This is because the focus of science education has leaned more towards the material content of science rather than the actual process of scientific inquiry [6]. Every time they solve a problem, some students go directly to the equations and mathematical processes without paying attention to the stages and steps of solving the problem. In addition, Astuti *et al.* [7] students have difficulty

ICWSDGs 2024, September 09-10, Banjarmasin, Indonesia

DOI 10.4108/eai.9-9-2024.2359068

analyzing questions due to the tendency of students to memorize answers from sample questions given by the teacher so students are confused in solving problems.

Based on the results of initial observations in the form of interviews with physics teachers in several Senior High Schools in Banjarmasin, it was obtained that, generally, physics teachers already know about problem-solving abilities. They said that the test instruments used already used problem-solving ability test instruments, but no analysis of question items had been carried out. In addition, the level of questions given starts from the cognitive level C1-C3.

Test instruments are one way to estimate the level of human ability indirectly through a person's response to several stimuli or questions [8]. The form of questions used in this study is a question with an essay form. There are advantages to tests using the form of essay questions compared to multiple-choice, that is, the form of multiple choice is influenced by guessing factors that can make the measurement accuracy results decrease. As a result, many students can answer physics problems correctly but do not know the physics concepts related to the problem. In addition, the multiple-choice test does not contain much information because it only brings up the student's final answer. With the use of essay question forms, it is considered to be able to measure the complex cognitive level of learners where they must be able to compile, interpret and unify knowledge, and use information in solving new problems. The quality of an instrument can be seen from its validity, reliability, level of difficulty, and separation index [9], [10], [11].

Validity is the accuracy of the interpretation obtained from the assessment results. In addition, validity means the extent of the accuracy and accuracy of a measuring instrument. The validation process involves gathering evidence to demonstrate the scientific basis for interpreting the score as planned. A test is said to be valid if it measures what will be measured. The test has high validity if the results match the criteria. In the sense of having parallels between tests and criteria [12].

Reliability means the consistency or stability of assessment results. The ability of a student is said to be reliable if measurements are made, the measurement results will be the same information, even though the examiners are different, the proofreaders are different or the question points are different but have the same characteristics [13].

The difficulty level of a question item is a parameter that can be used to determine how good the quality of a question item is. The difficulty level of a question item is the percentage or proportion of the test taker to answer correctly a question item. The high and low level of difficulty of a question item can be caused by the complexity of the subject matter and the conditions of the answer choices provided [13], [14].

The separation index of questions is the ability of a question item to distinguish students who have mastered the material asked and students who have not mastered the material asked. To analyze the separation index of students' questions, they are divided into 2 groups, namely the upper group and the lower group [12], [15]. The higher or greater the separation index of the question, the greater the question can distinguish between the high group and the low group [14].

In this study, the problem-solving indicators used are Heller's problem-solving indicators, namely describing problems, describing problems in physics concepts, planning solutions, implementing plans, checking, and evaluating [16].

The material chosen in the development of this problem-solving ability instrument is work and energy material because this material has a lot to do with phenomena in everyday life [17]. So the problem formulation is obtained, namely, "How is the quality of the instrument test the ability to solve problems in work and energy materials?". The purpose of this research is to produce a problem-solving ability test instrument on work and energy materials.

2 Method

2.1 Development Model

The type of research used is development research or Research and Development (R&D). This research develops cognitive instruments in the form of problem-solving abilities using the Borg and Gall adaptation development model, the steps are as in figure 1.

The research was conducted in May-June in the even semester of the 2021/2022 academic year in 3 schools in Banjarmasin. The subject of this study is a test instrument of the problem-solving ability of matter, work and energy. The test subjects were 23 class X students and 176 people for usage trials.

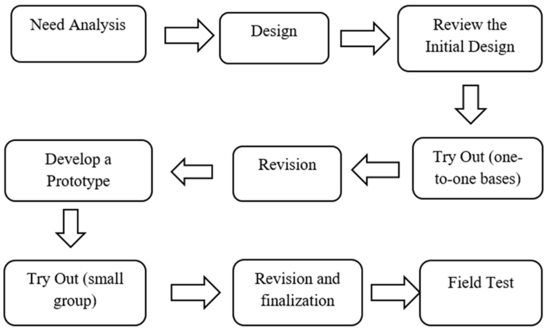


Fig. 1. Research Steps of the Borg and Gall Adaptation Model
Source: Sugiyono [18]

The steps taken in this study are:

a. Need analysis
Based on the results of research in previous studies, the development of problem-solving ability test instruments analyzed by the question items is still minimal. Therefore, this research is needed to produce a feasible problem-solving ability test instrument on work and energy materials.

b. Design
An interview was conducted with the physics teacher, then the results of the interview were analyzed. The result of the data analysis is that schools have applied problem-solving skills, but each question item has not been analyzed.

c. Review the initial design
At the product design stage, researchers designed the product developed, namely a test instrument in the form of a description of problem-solving skills in work and energy materials with as many as 17 questions. The test instrument is compiled, containing story questions accompanied by pictures. In addition, the product design also contains *a cover*, preface, table of contents, question grid, and problem-solving ability test questions.

d. Design Validation (one to one bases)
Furthermore, design validation was carried out to determine the quality of the problem-solving ability instrument that was prepared before being tested in the field. At the design validation stage, it was carried out by validators, namely 4 lecturers and 2 physics teachers.

e. Revision
   Test instruments that have been validated are revised based on suggestions from validators.
f. Develop a prototype
   The test instruments that were compiled were repaired before being tested to try out on small group.
g. Try out (small group)
   The product trial phase was carried out on a small scale, namely in class X MIPA students totaling 29 students. The number of questions given was 17 questions.
h. Revision and finalization
   Test instruments that have been tested are revised again and questions are received for trial use. There are 16 questions that accommodated for final instruments. The development of test instruments using the Rasch model can detect student misconceptions about test question items with the results of data analysis showing a match between students and the test instruments used [19].
i. Field test
   In the trial phase, the resulting product was tested on a large scale in three schools in Banjarmasin with a total of 176 students. The results of the data are then analyzed using the Rasch program on *Software* Winstep version 4.5.5 to find out the value of validity, reliability, difficulty level, and separation index of the problem. Product revisions were re-made to improve test instruments developed according to conditions in the field.

2.2 Data Analysis Techniques

2.2.1 Design Validation Analysis

This analysis was performed using Aiken's V formula adjusted to the criteria of $0.6 \leq V \leq 1.00$ in the valid category and $V < 0.6$ in the invalid category. The equation used is as follows [20]:

$$V = \frac{\sum s}{[n(c-1)]} \qquad (1)$$

$$s = r - l_0 \qquad (2)$$

The reliability value is calculated using *the Cronbach Alpha* formula in equations (4) and (5).

2.2.2 Product Trial Analysis

A. Validity of the question
   Correlation formula *Product Moment*:

$$r_{XY} = \frac{N(\sum XY) - (\sum X)(\sum Y)}{\sqrt{\left\{N \sum X^2 - (\sum X)^2\right\}\left\{N \sum Y^2 - (\sum Y)^2\right\}}} \qquad (3)$$

Information:
r_{XY}= Product moment *correlation coefficient*
N = total number of all students
X = each student's score on *the* item
$\sum$X = sum of all scores
Y = total score of each student
$\sum$X = total score of all students

Then the calculation results are compared with the following coefficient criteria.

Table 1. Question Validity Criteria

No	Value Validity	Criterion
1	$r_{xy} \leq 0{,}00$	Invalid
2	$0{,}00 < r_{xy} \leq 0{,}20$	Very low
3	$0{,}20 < r_{xy} \leq 0{,}40$	Low
4	$0{,}40 < r_{xy} \leq 0{,}60$	Enough
5	$0{,}60 < r_{xy} \leq 0{,}80$	Tall
6	$0{,}80 < r_{xy} \leq 1{,}00$	Very high

Source: Ratumanan [21]

B. Reliability of the problem

The reliability analysis of the problem using the formula *Alpha Cronbach* and the criteria of reliability are as follows:

$$r_{11} = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2}\right) \qquad (4)$$

$$\sigma_i^2 = \frac{\sum x_i^2 \left(\frac{\sum x_i^2}{N}\right)}{N} \; and \; \sigma_t^2 = \frac{\sum x_t^2 \left(\frac{\sum x_t^2}{N}\right)}{N} \qquad (5)$$

Information:
r_{11} = Reliability of the problem
n = number of question items (items)
$\sum \sigma_i^2$ = sum of score variances per *item*
σ_t^2 = total variance
$\sum x_i^2$ = Number of scores per question *item* number
$\sum x_t^2$ = total score of questions
N = number of students

The criteria for the reliability of the question items are as follows:

Table 2. Question Reliability Criteria

No	The Value of Reliability	Criterion
1	$0{,}0 \leq r_{11} < 0{,}2$	Very low
2	$0{,}2 \leq r_{11} < 0{,}4$	Low
3	$0{,}4 \leq r_{11} < 0{,}6$	Enough
4	$0{,}6 \leq r_{11} < 0{,}8$	Tall
5	$0{,}8 \leq r_{11} \leq 1{,}0$	Very high

Source: Meilanie [22]

C. Difficulty level

Determining the level of difficulty in the form of a description question can be used formula:

$$TK = \frac{Average\ Score}{Max\ Score} \quad (6)$$

The criteria for the difficulty index of the questions are as follows:

Table 3. Difficulty Criteria

No	Difficulty Index	Criterion
1	0,00 – 0,30	Difficult
2	0,31 – 0,70	Keep
3	0,71 – 1,00	Easy

Source: Matondang *et al* [23]

D. Separation Index

The formula used to determine the separation index of the question is as follows:

$$DP = \frac{Mean\ Top\ Category - Mean\ Bottom\ Category}{Max\ Score} \quad (7)$$

Table 4. Criteria for Discriminating Power of Questions

No	Differentiation Power Index	Criterion
1	$DP \leq 0{,}0$	Very ugly
2	$0{,}0 < DP \leq 0{,}20$	Ugly
3	$0{,}20 < DP \leq 0{,}40$	Enough
4	$0{,}40 < DP \leq 0{,}70$	Good
5	$0{,}70 < DP \leq 1{,}00$	Excellent

Source: Meilanie [22]

2.2.3 Usage Trial Analysis

Question items that have a good match occur if the value of *Outfit Mean Square* (0.5<MNSQ<1.5), *Outfit Z-standard* (-2.0<ZSTD<+2.0) and *Point Measure Correlation* (0.4<Pt Mean Corr<0.85). A problem is said to contain bias if the probability value is less than 5% (<0.05). As for the reliability of the questions in the Rasch program, it is known based on the results of the output table summary statistics by looking at the value of *Alpha Cronbach*.

3 Results

3.1 Design Validation and Revision

Table 5. Design Validation Analysis Results

No	Aspects	Average	Criterion
1	Material	0,78	Valid
2	Construction	0,68	Valid
3	Language	0,63	Valid
4	Troubleshooting Capabilities	0,77	Valid
	Reliability	0,98	Very High

3.2 Product Trials and Product Revisions

This product was tested on a small scale on class X MIPA students who had studied work and energy materials with a total of 29 students. The problem-solving ability test instrument used consists of 17 description questions.

A. Validity

Table 6. Product Trial Validity Results

No	Question Number	Criterion	
1	11	Very high	
2	2, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17	High	
3	1	Moderate	
4	3	Very low	
Valid	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17	16 Questions	90 %
Invalid	3	1 Question	10

B. Reliability

Table 7. Product Trial Reliability Results

N	r_{Table}	Reliability	Criterion
29	0,367	0,92	Very high

C. Difficulty level

The results of the difficulty level analysis are briefly presented in the following table.

Table 8. Product Test Difficulty Results

No	Question Number	Criterion	Percentage
1	1, 8, 9, 12, 13, 14, 15, 16, 17	Difficult	52,94 %
2	2, 4, 5, 6, 7, 10, 11	Quite	41,18 %
3	3	Easy	5,88 %

D. Separation index

Table 9. Product Trial Separation Index Results

No	Question Number	Criterion	Percentage
1	7, 10, 11	Good	17,65 %
2	2, 4, 5, 6, 8, 9, 12, 14	Moderate	47,06 %
3	1, 13, 15, 16, 17	Bad	29,41 %
4	3	Very bad	5,88 %

3.3 Usage Trial

The following are the results of the trial analysis using the Rasch program:

A. Validity

Table 10. Results of Field tes Validity

Entry Number	*Outfit*		*Pt-Measure Correlation*	*Probability*
	MNSQ	**ZSTD**		
1	0,39	-7,65	0,35	0,0000
2	1,34	3,26	0,28	0,0000
3	1,62	4,15	0,38	0,0000
4	0,54	-5,57	0,69	0,0900
5	0,85	-1,21	0,52	0,0000
6	1,05	0,26	0,27	0,0375
7	0,68	-2,38	0,53	0,0038
8	0,60	-1,65	0,55	0,0000
9	0,50	-1,76	0,60	0,0243
10	1,53	2,91	0,35	0,9725
11	0,55	-1,20	0,48	0,1109
12	0,23	-1,61	0,40	0,7710
13	0,88	-0,36	0,49	0,1634
14	0,20	-1,74	0,42	0,2156
15	1,17	0,69	0,38	0,0001
16	1,39	0,95	0,19	0,3462

B. Reliability

In the Rasch program, reliability can be known based on the results *Output Table* 3.1

Summary statistics by looking at the value of *Alpha Cronbach*, *Separation Person*and *Separation Item*.

Table 11. Reliability Results of Usage Trials

Name	Information
Alpha Cronbach	0,69
Separation Person	1,12
Item Separation	5,25

C. Difficulty level

Table 12. Field Test Difficulty Level Results

No	Question Number	Criterion	Percentage
1	12, 14	Difficult	12, 50 %
2	3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 15, 16	Quite	75,00 %
3	1, 2	Easy	12, 50 %

D. Separation index

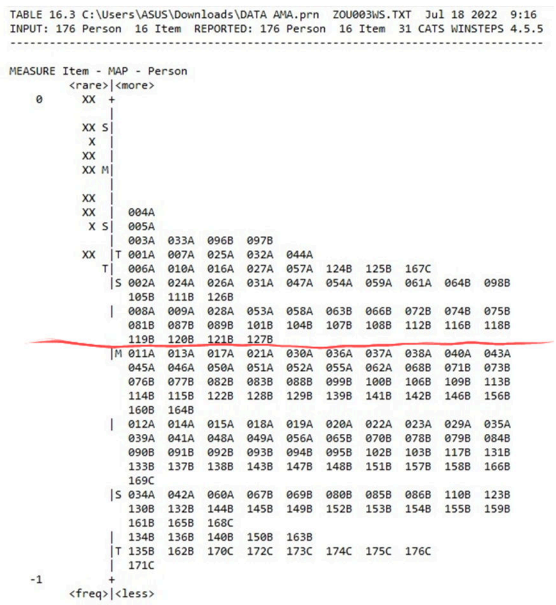
Based on the results of the data, the separation index value of each question item is obtained with the following categories.

Table 13. Separation Index Results of Discharging Trials

No	Question Number	Criterion	Percentage
1	1, 2, 4	Good	18,75 %
2	3, 5, 7, 8, 10, 15	Moderate	37,50 %
3	6, 9, 11, 13, 16	Bad	31,25 %
4	12, 14	Very bad	12,50 %

E. Student capability

In addition to analyzing the question items, it is also necessary to analyze the ability of students to work on the questions.



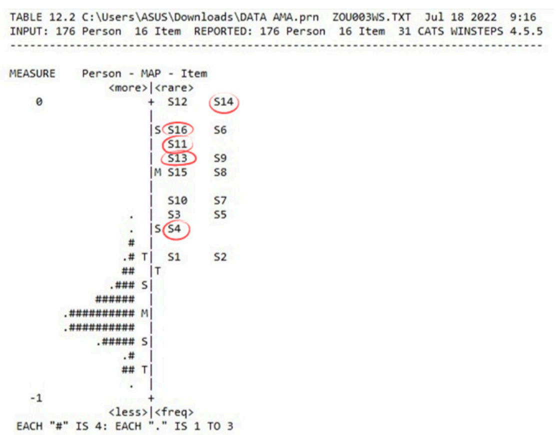
```
TABLE 16.3 C:\Users\ASUS\Downloads\DATA AMA.prn  ZOU003WS.TXT  Jul 18 2022  9:16
INPUT: 176 Person  16 Item  REPORTED: 176 Person  16 Item  31 CATS WINSTEPS 4.5.5
--------------------------------------------------------------------------------

MEASURE Item - MAP - Person
        <rare>|<more>
   0      XX  +
              |
          XX S|
           X  |
          XX  |
          XX M|
              |
          XX  |
          XX  |  004A
           X S|  005A
              |  003A  033A  096B  097B
          XX  |T 001A  007A  025A  032A  044A
             T|  006A  010A  016A  027A  057A  124B  125B  167C
              |S 002A  024A  026A  031A  047A  054A  059A  061A  064B  098B
                 105B  111B  126B
              |  008A  009A  028A  053A  058A  063B  066B  072B  074B  075B
                 081B  087B  089B  101B  104B  107B  108B  112B  116B  118B
                 119B  120B  121B  127B
              |M 011A  013A  017A  021A  030A  036A  037A  038A  040A  043A
                 045A  046A  050A  051A  052A  055A  062A  068B  071B  073B
                 076B  077B  082B  083B  088B  099B  100B  106B  109B  113B
                 114B  115B  122B  128B  129B  139B  141B  142B  146B  156B
                 160B  164B
              |  012A  014A  015A  018A  019A  020A  022A  023A  029A  035A
                 039A  041A  048A  049A  056A  065B  070B  078B  079B  084B
                 090B  091B  092B  093B  094B  095B  102B  103B  117B  131B
                 133B  137B  138B  143B  147B  148B  151B  157B  158B  166B
                 169C
              |S 034A  042A  060A  067B  069B  080B  085B  086B  110B  123B
                 130B  132B  144B  145B  149B  152B  153B  154B  155B  159B
                 161B  165B  168C
              |  134B  136B  140B  150B  163B
              |T 135B  162B  170C  172C  173C  174C  175C  176C
              |  171C
  -1          +
        <freq>|<less>
```

Fig. 2. Distribution Map of Student Ability Levels

F. Instrument analysis of problem-solving ability tests

In general, the questions in the test are more difficult when compared to the ability of students.



```
TABLE 12.2 C:\Users\ASUS\Downloads\DATA AMA.prn  ZOU003WS.TXT  Jul 18 2022  9:16
INPUT: 176 Person  16 Item  REPORTED: 176 Person  16 Item  31 CATS WINSTEPS 4.5.5
--------------------------------------------------------------------------------

MEASURE    Person - MAP - Item
                <more>|<rare>
   0                  +  S12    S14
                      |
                      |S S16    S6
                      |  S11
                      |  S13    S9
                      |M S15    S8
                      |
                      |  S10    S7
                    . |  S3     S5
                    . |S S4
                    # |
                 .# T|  S1     S2
                   ## |T
               .### S|
               ###### |
          .########## M|
          .##########  |
              .##### S|
                   .# |
                  ## T|
                    . |
  -1                  +
                <less>|<freq>
 EACH "#" IS 4: EACH "." IS 1 TO 3
```

Fig. 3. Item *Variable Maps*

4 Discussion

4.1 Instrument Quality of Design Validation

The test instrument prepared is a problem-solving ability test instrument on work and energy material for class X students in even semesters. This instrument was validated by 6 validators, namely 4 academics (lecturers) and 2 practitioners (physics teachers). The aspects assessed are material, construction, language, and problem-solving skills. In line with Amelia *et al.* [24], before the test instrument is used, it is necessary to review the assessing instrument. An authentic assessment offers clearer insights into learners' abilities [25]. Factors that need to be observed in qualitative question reviews are language, construction, material, content aspects, scoring guidelines, and answer keys. Based on Table 5 of the validation results from the six validators, it was found that the developed instrument design was included in the valid category. This is supported based on the opinion of Adawiyah & Wisudawati [26] who state that a test instrument is said to be appropriate if the assessment of all aspects, namely aspects of topics, construction, and grammar by experts is good. However, the instrument needs to be improved based on the suggestions given. In addition to validity, instruments are also seen in terms of reliability. The reliability results obtained based on the assessment by validators are included in the very high category, thus showing that the test instrument has good reliability.

4.2 Instrument Quality from Product Trials

A. Validity

Based on the analysis of product trials, namely in Table 6, it was found that there were 16 valid questions and 1 invalid question. Invalid questions cannot be used in usage trials. A test is said to be valid if the test is under real circumstances [27]. A test or measuring instrument is said to have high validity if the instrument performs its measuring function, or provides measuring results that follow the purpose for which the measurement is made. This means that the measurement result of the measurement is a quantity that accurately reflects the facts or circumstances of what is measured [28]. Test instruments that have invalid measuring power will not be able to provide information regarding the ability of learners. In line with statements from Pratama [29] and Purwanto [30], a test must be valid so that it can measure what it wants to measure. While questions with high validity can be used, and questions with medium and low validity can also be used, the questions need to be improved.

B. Reliability

Reliability in classical theory is determined by the formula *Alpha Cronbach* indicating internal consistency [31], based on Table 7 so that the results obtained fall under the very high criteria. These results show that the instruments developed have been reliable. In line with the opinion of Wati & Mahtari [32], If the test is said to be reliable, then the instrument can be used continuously to make measurements because of its good qualifications.

C. Difficulty level

The level of difficulty of a question item is one of the parameters of the question item that is very useful in analyzing a test. Based on the results of research in Table 8, the question items are relatively easy because students can answer the question correctly. While the question items are classified as difficult because students do not answer the questions or the student's answers are wrong. Following the statement by Fatimah and Alfath [33], a test should not be too simple, nor should it be too difficult.

D. Separation index

In product trials, it was found that there was 1 question item with very bad separation index, which can be seen in Table 9. The follow-up of the results of the separation index analysis of this question is to abort or discard questions with very bad criteria as much as 1 question. Questions that have low discriminating power are caused by too difficult questions or problems containing bias [15]. Thus, the separation index of question items becomes a benchmark in determining questions that cannot be used, can be used with revision, and questions without revision.

4.3 Instrument Quality from Trial Use

A. Validity

Determining the validity or absence of a question using the Rasch model can be seen based on the values of MNSQ, ZSTD, and *Point Measure Correlation* with criteria: (a) value *Outfit Mean Square* (MNSQ) a good one is $0.5<\text{MNSQ}<1.5$; (b) value *Z-standard output* (ZSTD) accepted is $-2<\text{ZSTD}<+2$; and (c) value *Point Measure Correlation (Pt Mean Corr)* received is $0.4<\text{Pt Mean Corr}<0.85$ [34].

Sumintono and Widhiarso [34] state that a biased question item will have a probability value of $<5\%$. Based on the results in Table 10, question items that have a probability value below 0.05 questions will be discarded. So, the total questions that are discarded and cannot be used are based on question items that do not *fit* where it does not meet the criteria of MNSQ, ZSTD, and *Pt Mean Corr*. The bias question items are questions number 1, 2, 3, 5, 6, 7, 8, 9, 10 and 15.

B. Reliability

Based on Table 11, values *Alpha Cronbach* which is used to measure the interaction between respondents and question items (*Items*) shows results with sufficient category, it can be stated that the test instrument is reliable. This is according to research Puspasari and Puspita [35] which states that *Alpha Cronbach* is a measure of reliability that has values ranging from zero to one. When a variable shows a value *Alpha Cronbach>0.60 then it can be concluded that the variable can be said to be reliable or consistent in measuring.* The factors that affect the reliability value are the greater the number of respondents, the greater the accuracy, the longer the test time, the greater the accuracy, the narrower the range of difficulty levels of the question items, the greater the reliability, and the misinterpretation of the question items [36].

C. Difficulty level

The results of the calculation of the difficulty level in the usage trial are shown in Table 12, that the further down, the easier the difficulty of the question. This is according to Sumintono &; Widhiarso [34] that is, the degree of difficulty of the question item (*Item Measure*) provides a value display *logit* from each question item in detail.

D. Separation index

The higher or greater the separation index of the question, the greater the question can distinguish between the high group and the low group. In this study, it can be seen that based on Table 13 many questions have obtained separation index with bad criteria. However, question items with poor separation index should not be discarded. To determine whether or not the question items are discarded can be seen by comparing the results of the separation index and [14] *Item Variable Maps* (Figure 3).

E. Student capability

Analysis of the ability of students is carried out to see their ability to solve problems. Based on Figure 2, students who have high abilities have the code 004A. These students have the highest ability because they get the highest scores compared to other students. As for students who have low abilities, it has a code of 171C. The barrier depicted shows that students are divided into 2 groups with the symbol "M" being the middle boundary. To find out the grouping of students can be known by using values *separation*. By using equations $\mathrm{H} = \{(4 \times Separation) + 1\}/3$ and value *separation person* as much as 1.12 (Table 11), obtained a value of 1.827 (rounded to 2) which means that this instrument groups students into 2 groups.

F. Instrument analysis of problem-solving ability tests

Based on Figure 3, it is found that the most difficult problem-solving ability test instruments are found in questions number 12 and number 14. From the answers of learners based on problem-solving indicators, that on each indicator obtained low results. This shows that students have not been able to understand the problem and describe it. In addition, some are also not correct in entering known data into the formula and there are still many students who are not correct or do not write the conclusion.

The reason many students do not answer the question is because students do not finish doing the question because of time constraints and students do not know the answer to the question [37]. The results of the research show that the students' ability to solve physics problems on work and energy is still low. Mustofa & Rusdiana [38], problem-solving is a process, namely the process of thinking or the process of applying knowledge that has been obtained. On the other hand, Tongli *et al.,* [39] stated that problem solving is being able to demonstrate knowledge, make the right decision, make solution, and cultivate a clear mind to solve the problem. This problem-solving ability is a concern for teachers so that they are always trained in the learning process. With frequent physics problem-solving exercises, it is expected that students' problem-solving abilities will increase. In addition, according to Yusrizal *et al.*, [40]. The main purpose of asking questions to students is to find out whether the material taught has been mastered by students completely or not.

Based on the data analysis conducted, these results are supported by the results of research conducted by Taqwa *et al*. [41], namely the test instrument as many as 15 questions are valid and reliable. The results of the study can also be compared with research conducted by Wartono *et al.* [42], which obtained 2 out of 3 points of description questions that were developed as valid, have high reliability, and moderate level of difficulty of the questions. Based on these two relevant studies, it can be concluded that the developed instrument is suitable for use, but the bias question needs to be discarded so that the final stage problem-solving ability test instrument totaling 5 description questions can be used as a test of student learning outcomes.

5 Conclusions

Based on the results and discussion of the research that has been done, it can be concluded that the problem-solving ability test instrument developed based on the results of design validation, product trials, and usage tests (rasch model) is valid and has good reliability. The difficulty level of the developed problem-solving ability test instrument has a medium category for all questions. The separation index of the problem-solving ability test instrument developed consists of 16 questions, 5 acceptable questions, and 11 discarded questions including invalid questions and biased questions.

References

[1] A. R. Sinensis, H. Firman, I. Hamidah, and Muslim, “Pengembangan instrumen tes termodinamika untuk mengukur kemampuan pemecahan masalah (PsACAr) pada mahasiswa calon guru fisika,” *Jurnal Inovasi dan Pembelajaran Fisika*, pp. 122–129, 2019.

[2] H. Ulya, R. B. Rudibyani, and T. Efkar, “Pengembangan modul kimia berbasis problem solving pada materi asam basa arrbenius,” *Jurnal Pendidikan dan Pembelajaran KImia*, vol. 7, no. 1, pp. 129–141, 2018.

[3] A. Darmawan, B. N. Asa, F. Kurniawan, and R. Nukhba, “Pengembangan instrumen tes pemecahan masalah bagi mahasiswa jurusan fisika pada materi dinamika partikel,” *Jurnal Pendidikan Fisika dan Keilmuan*, pp. 55–64, 2020.

[4] H. W. Qamariah and Nurkamilia, “Development of physics module based problem based learning for electrical engineering students,” AIP Conference Proceedings, 2024.

[5] A. Y. Arief, *Profil kemampuan pemecahan masalah fisika siswa kelas X MAN Yogyakarta 1*. Yogyakarta: UIN Sunan Kalijaga Program Studi Pendidikan Fisika, 2017.

[6] F. Ahmad, Mastuang, and Misbah, “The implementation of guided discovery learning model to improve students’ characters of responsibility and academic skills,” *Journal of Physics: Conference Series*, vol. 2104, 2021.

[7] N. H. Astuti, A. Rusilowati, B. Subali, and P. Marwoto, “Analisis kemampuan pemecahan masalah model polya materi getaran, getaran, gelombang,dan bunyi siswa smp,” *Unnes Physics Education Journal*, pp. 1–8, 2020.

[8] C. F. Pasani, R. F. Putri, and R. Yulinda, “Feasibility of learning instrument with sainti-ka (characteristic of scientific) of Junior High Scholl on environmental pollution topic,” *Journal of Physics: Conference Series*, vol. 1832, 2021.

[9] D. Mardapi, *Teknik penyusunan instrumen tes dan nontes*. Jogjakarta: Mitra Cendikia, 2008.

[10] Aristiawan and I. E, “Developing instrument of essay test to measure the problem-solving skill in physics,” *Jurnal Pendidikan Fisika Indonesia*, vol. 16, no. 2, pp. 72–82, 2020.

[11] M. I. Ismail, *Asesmen dan evaluasi pembelajaran*. Makassar: Cendekia Publisher, 2020.

[12] Z. Arifin, “Kriteria Instrumen dalam suatu Penelitian,” *Jurnal THEOREMS (The Original Research of Mathematics)*, vol. 2, no. 1, pp. 28–36, 2017.

[13] H. Retnawati, *Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, dan Psikometri)*. Yogyakarta: Parama Publishing, 2016. [Online]. Available: www.nuhamedika.gu.ma

[14] N. Hanifah, “Perbandingan tingkat kesukaran, daya pembeda butir soal dan reliabilitas tes bentuk pilihan ganda biasa dan pilihan ganda asosiasi mata pelajaran ekonomi,” *Jurnal Program Studi Bimbingan Konseling Universitas Indraprasta PGRI Jakarta*, vol. 6, no. 1, pp. 41–55, 2014.

[15] D. R. Sari, S. Wahyuni, and R. W. Bachtiar, “Pengembangan instrumen tes multiple choice high order thinking pada pembelajaran fisika berbasis e-learning di sma,” *Jurnal Pembelajaran Fisika*, pp. 100–107, 2018.

[16] P. Heller, R. Keith, and S. Anderson, “Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving,” *American Journal of Physics*, vol. 60, no. 7, pp. 627–636, 1992.

[17] K. Lestari, H. T. Maria, and D. Mahmuda, “Penerapan penyelesaian masalah heller untuk meningkatkan kemampuan menyelesaikan soal materi gerak lurus,” *Jurnal Program Studi Pendidikan FIsika FKIP Untan Pontianak*, pp. 1–9, 2019.

[18] Sugiyono, *Metode penelitian kuantitatif, kualitatif dan R&D*. Bandung: Alfabeta, 2013.

[19] R. P. Yudha, “Higher Order Thinking Skills (HOTS) test instrument: validity and reliability analysis with the rasch model,” *Mathematics Education Learning and Teaching*, vol. 12, no. 1, pp. 21–38, 2023.

[20] N. Fadieny and A. Fauzi, “Validitas e-modul fisika terintegrasi materi bencana petir berbasis experiential learning,” *Jurnal Penelitian dan Pembelajaran Fisika*, vol. 7, no. 1, pp. 17–25, 2021.

[21] T. G. Ratumanan, “Pengaruh modul pembelajaran dan gaya kognitif terhadap hasil belajar matematika siswa sltp di kota ambon,” *Jurnal Pendidikan Dasar*, vol. 5, no. 1, pp. 35–39, 2003.

[22] E. P. Meilanie, *Pengembangan*. Semarang: Universitas Negeri Semarang, 2020.

[23] Z. Matondang, E. S. Djulia, and J. Simarmata, *Evaluasi hasil belajar*. Medan: Yayasan Kita Menulis, 2019.

[24] H. Amelia, M. Wati, and S. Hartini, “Pengembangan instrumen kognitif untuk mengukur penalaran siswa SMP di Kota Banjarmasin pada materi cahaya,” *Berkala Ilmiah Pendidikan Fisika*, vol. 6, no. 1, pp. 83–97, 2018.

[25] T. Khairunnisa, Supramono, and A. Sholahuddin, “THE USE OF PORTFOLIO INSTRUMENTS TO ASSESS SCIENCE PROCESS SKILLS OF STUDENTS IN JUNIOR HIGH SCHOOL,” *European Journal of Education Studies*, pp. 173–181, 2019.

[26] R. Adawiyah, “Pengembangan Instrumen Tes Berbasis Literasi Sains Dalam Aspek Menjelaskan Fenomena Ilmiah Materi Pokok Energi,” 2017.

[27] M. Wati and S. Miriam, “Pengembangan instrumen kognitif literasi sains pada pokok bahasan tekanan di kelas VII SMP kota Banjarmasin,” *Jurnal Ilmiah Pendidikan Fisika*, vol. 1, no. 3, pp. 113–125, 2017.

[28] R. U. Sudaryono, Q. Aini, Y. I. Graha, and N. Lutfiani, “Validity of test instruments,” *Journal of Physics: Conference Series*, pp. 1–11, 2018.

[29] D. Pratama, “Analisis kualitas tes buatan guru melalui pendekatan item response theory (irt) model rasch,” *Tarbawy: Jurnal Pendidikan Islam*, vol. 7, no. 1, pp. 61–70, 2020.

[30] Purwanto, *Evaluasi hasil belajar*. Yogyakarta: Pustaka Belajar, 2016.

[31] H. Hairida *et al.*, “Evaluating Digital Literacy of Pre-service Chemistry Teachers: Multidimensional Rasch Analysis,” *Journal of Science Education and Technology*, vol. 32, pp. 1–12, Aug. 2023, doi: 10.1007/s10956-023-10070-z.

[32] M. Wati and S. Mahtari, “Pengembangan instrumen kognitif fisika siswa SMP,” *Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: e-Saintika*, vol. 1, no. 1, pp. 45–56, 2017.

[33] L. U. Fatimah and K. Alfath, “Analisis kesukaran soal, daya pembeda dan fungsi distraktor,” *Jurnal Komunikasi dan Pendidikan Islam*, vol. 8, no. 2, pp. 37–64, 2019.

[34] B. Sumintono and W. Widhiarso, *Aplikasi pemodelan rasch pada assessment pendidikan*. Bandung: Trim Komunikata, 2015.

[35] H. Puspasari and W. Puspita, “Uji validitas dan reliabilitas instrumen penelitian tingkat pengetahuan dan sikap mahasiswa terhadap pemilihan suplemen kesehatan dalam menghadapi covid-19,” *Jurnal Kesehatan*, vol. 13, no. 1, pp. 65–71, 2022.

[36] A. Kadir, “Menyusun dan menganalisis tes hasil belajar,” *Jurnal Al-Ta’dib*, vol. 8, no. 2, pp. 70–81, 2015.

[37] L. Rohmah, S. Handono, and Yushardi, “Analisis kesalahan siswa dalam memecahkan masalah fisika berdasarkan polya pada pokok bahasan fluida statis di SMAN Jember,” *Jurnal Pembelajaran Fisika*, vol. 7, no. 4, pp. 328–333, 2018.

[38] H. Mustofa and D. Rusdiana, “Profil kemampuan pemecahan masalah siswa pada pembelajaran gerak lurus,” *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, vol. 2, no. 2, pp. 15–22, 2016.

[39] B. Tongli, M. S. Zurweni, M. A. Tuasikal, A. D. Harnoko, A. Rozak, and Mulyono, “Investigating the correlation digital literacy, instructional leadership, and intelligence on work performance,” *Journal of Infrastructure, Policy and Development*, vol. 6356, 2024.

[40] Yusrizal, Suliyanah, and T. H. Basri, “Analysis of knowledge, understanding, and skills of physics teachers of state senior high schools in developing and analyzing test items,” *Jurnal Pendidikan IPA Indonesia*, vol. 6, no. 2, pp. 335–340, 2017.

[41] A. Taqwa, E. Purwaningsih, and Sulur, “Pengembangan instrumen tes kemampuan pemecahan masalah mahasiswa pada topik usaha dan energi,” *Jurnal Penelitian Pembelajaran Fisika*, vol. 11, no. 2, pp. 149–156, 2020.

[42] B. Wartono, R. S. J, and D. I. Rahmatina, “Development of problem solving ability test on the work-energy material,” *Journal of Education and Learning (EduLearn)*, vol. 13, no. 2, pp. 268–276, 2019.