P-ISSN 2746-9034 || E-ISSN 2774-5066
Indonesian Physics Education Research
Volume 03, No.01, June 2022, pp. 44-53

QUALIFICATION OF PHYSICS TEST INSTRUMENTS LIKE-PISA

Khairina Rangkuti[1], Wawan Bunawan[2]
[1,2] Universitas Negeri Medan, Medan 2021, Indonesia
Email: khairinarangkuti@gmail.com

Abstract

This research aims to develop a physics test instrument based on scientific literacy PISA (Program for International Student Assessment). Development of test instruments using 4D models (Define, Design, Development, Disseminate). The test instrument was tested by expert validity with a CVI score of 1. The test instrument tested involved 130 participants. The results of the item validity test showed that the 20 items tested were valid with a Cronbach Alpha reliability value of 0.816 ("very good" category). The discrimination power test results showed that the developed items are in the "good" and "very good" categories. The results of the difficulty level test showed that there was one difficult item, 11 medium questions, and eight easy questions. This research shows that the physics science literacy skills of students who participate in this research show that students achieve good scores on average. Students who participated in this research had the highest achievement in the competence of evaluation and design of scientific investigations and had the same achievement in both the content and the procedural knowledge aspects.

Keywords: Test Instrument, Physics, Scientific Literacy, PISA

Indonesia has participated in the PISA (Programme of International Students Assessment) program for seven rounds since 2000. PISA is a three-year program held by the Organization for Economic Operation and Development to assess the ability of 15-year-old students in mathematical, scientific, and financial literacy (OECD, 2019). The 2018 PISA assessment announced by the OECD in 2019 showed that Indonesia was in the 10th lowest rank of the 79 participating countries. The 2018 PISA results also show that Indonesian students' average scientific literacy ability is still below the average ability of ASEAN students and countries with similar socioeconomic conditions, such as Peru and Brazil (Kemendikbud, 2019). The results of PISA 2018 align with the research (Permatasari & Fitriza, 2019), which shows that high school students have low scientific literacy skills in content, context, and competence. The low ability of students' scientific literacy in these three aspects is because students have a low ability to identify scientific opinions, effectively do literature searches, understand experiment design, and convert data from one form to another (Rizkita et al, 2016). This is because students are unfamiliar with working on questions that test students analytical skills. The test instruments used in evaluation in schools use questions that only test students' ability to remember the lessons they have learned (Sutrisna, 2021). The results of observations and interviews conducted at the MAN 1 Medan school show that MAN 1 Medan has not implemented scientific literacy in either learning or assessment. The Physics teacher interviewed also stated that he was unfamiliar with scientific literacy.

Based on the above background, the following problems can be formulated: (1) the low achievement and scientific literacy ability of Indonesian students at PISA 2018; (2) the Lack of development of scientific literacy instruments, especially in high school physics learning; (3) Students are less trained in working on scientific literacy-based questions. The purpose of this research is to develop a physics test instrument that is under the test qualifications and PISA qualifications. This research is expected to provide practical benefits: (1)Contribute to high school physics assessments; (2) Provide an overview of students' scientific literacy in learning physics. This research is also expected to provide theoretical benefits: (1) Contribute and provide input in evaluating high school physics education, especially PISA-based tests; (2) As a material for consideration and research study for the development of other PISA-based test instruments.

METHODS

The type of research used is research and development with a 4D research model (Define, Design, Develop, and Disseminate) by (Thiagarajan, 1974). This research involved 10th-grade students of MAN 1 Medan as research participants. The data in this study were collected using validation sheet instruments, questionnaires, and written tests. This research begins with a definition by conducting an initial analysis through literature studies and observations related to the use of the PISA scientific literacy test instrument, as well as curriculum analysis by analyzing core competence, basic competence, and passing standards in the curriculum 2013 on the topic of force, motion, work, and energy as the basis for determining the scientific indicators and blueprints of the developed test instruments. The design is carried out by determining the format of the questions, the material being tested, the preparation of the test instrument blueprint, and the preparation of the questions, then developed by validating the test instrument by an expert. After being stated valid by five experts, the test instrument was tested in small groups to determine validity, reliability, level of difficulty, and discriminating power. The results will be used to revise the test instrument to produce prototype III, which will be tested on a large scale to determine the validity, reliability, difficulty level, discrimination, student responses to questions, and students' scientific literacy level. Then the questions that have been tested are distributed (disseminated) by submitting the test instrument to the MAN 1 Medan.

The results of expert validation were analyzed using the research equation analyzed using the Content Validity Ratio Approach using the Lawshe equation (LAWSHE, 1975)

$$CVR = (ne - N/2)/(N/2) \quad (1)$$

Information :

Ne = number of experts who agree

N = number of experts who validate

The item is said to be valid if the CVR value is greater than the CVR critical value. Table 1 is a critical value of CVR.

Table 1 CVR Critical Value Table

Number of Validators	CVR Critical Value
5	0.736
6	0.672
7	0.622
8	0.582

After getting the CVR value, the CVI (Content Validity Index) value will be calculated using the formula:

$$CVI = \frac{CVR}{Number\ of\ Item\ Test} \quad (2)$$

The validation of the items tested on a small and large scale was analyzed using Pearson's moment correlation with the equation:

$$r_{XY} = \frac{N.\ \sum X.Y - (\sum X)(\sum Y)}{\sqrt{(N.\sum X^2 - (\sum X)^2).\ (N.\sum Y^2 - (\sum Y)^2)}} \quad (3)$$

Information :

$r$xy = correlation coefficient between variable X and variable Y

N = number of participants

X = student score on each item

Y = total score of each respondent

Table 2 Question validity criteria

Score	Criteria
0.80 < r_{xy} 1.00	Very high
0.60 < r_{xy} 0.80	High
0.40 < r_{xy} 0.60	Enough
0.20 < r_{xy} 0.40	Low
r_{xy} 0.20	Very low

The difficulty level of the questions will be analyzed by determining the index of the difficulty level of the questions based on a classical theory with the equation:

$$p = \frac{\Sigma B}{N} \tag{4}$$

Information:

p = the proportion of correct answers on certain items

B = number of participants who answered correctly

N = number of test participants who answered

Table 3 Level Classification Problem Difficulty

Index	Difficulty Level
0.00 – 0.30	Easy
0.31 – 0.70	Medium
0.71 – 1.00	Hard

The discriminatory power of the questions was analyzed using the biserial point correlation index with the equation:

$$r_{pbis} = \left[\frac{\overline{X_1} - \bar{X}}{s_X}\right]\sqrt{\frac{p_1}{1 - p_1}} \tag{5}$$

Information:

r pbis = biserial point coefficient

$\overline{X_1}$ = mean score of X participants who answered correctly

X 1 = mean score

Table 4 Differential Power Classification Question

Score	Criteria
0.40	Very good
Score	**Criteria**
0.30 – 0.39	Good
0.20 – 0.29	Minimum, Needs to be fixed
0.19	Not good

The reliability of the questions was analyzed using Cronbach's Alpha approach using the equation:

$$r_{11} = \left[\frac{n}{n-1}\right]\left[1 - \frac{\Sigma\sigma_i^{\,2}}{\sigma_r^{\,2}}\right] \tag{6}$$

Information:

r 11 = Instrument

n = number of questions

$\Sigma\sigma_i^{\,2}$ = total score variance for each item

$\sigma_r^{\,2}$ = Total score variance

The values obtained from calculations using the above equation are interpreted with the values in the table below:

Table 1. Reliability Criteria

Score	Information
0.90 < r_{11} 1.00	Very high
0.70 < r_{11} 0.90	tall
0.40 < r_{11} 0.70	enough
0.20 < r_{11} 0.40	low
r_{11} 0.20	Very low

The students responses obtained were analyzed using the following equation:

$$\%respones = \frac{score\ obtained}{maximum\ score} \times 100\% \tag{7}$$

Table 2. Instrument Eligibility Level Response-Based Test Student

Score	Eligibility Level
76% - 100%	Very good
51% - 75%	Well
26% - 50%	Enough
0% - 25%	Not good

RESULT AND DISCUSSION

This research produces a blueprint of questions, scoring guidelines, question sheets, and answers that are validated by experts and then tested on a small and large scale. At the expert validation stage, the items developed got an average ratio coefficient value of 1 with a validation coefficient index of 1. This shows that the test instrument developed is valid with very good criteria. After being declared valid by experts, a trial was carried out on a small scale and the results of the small-scale trial will be a consideration to revise the instrument test and then the revised instrument test will test in a large-scale trial.

The trial was conducted by testing the test instrument on 130 10th-grade students majoring in science. The results are listed in Table 7.

Table 3. Test Results

Item Question	r_{xy}	r_{pbis}	**Level Difficulty**
item 1	0.42	0.31	0.67
item 2	0.60	0.53	0.68
item 3	0.47	0.37	0.73
item 4	0.43	0.34	0.29
item 5	0.56	0.48	0.72
item 6	0.43	0.33	0.67
item 7	0.53	0.47	0.52
item 8	0.43	0.32	0.57
item 9	0.49	0.41	0.80
item 10	0.46	0.37	0.74
item 11	0.40	0.35	0.68
item 12	0.53	0.47	0.34
item 13	0.55	0.47	0.63
item 14	0.49	0.41	0.78
item 15	0.54	0.46	0.79
item 16	0.43	0.33	0.70
item 17	0.39	0.32	0.76
item 18	0.56	0.47	0.67
Item Question	r_{xy}	r_{pbis}	**Level Difficulty**
item 19	0.41	0.31	0.65
item 20	0.40	0.30	0.75

Based on the rXY value Table 7 shows 20 questions developed are valid empirically. The items' validity requires compatibility between the test instrument used in measuring the ability to be measured.

The reliability of the test instrument in large-scale trials has a Cronbach's Alpha value of 0.816, indicating that the test instrument developed is reliable following (Basuki, 2014) and that the test instrument is declared reliable if the Cronbach's Alpha value is between 0.60 and 1.

Discrimination power is the ability of questions to show differences between students with high and low abilities (Arikunto, 2012). The discrimination power is represented with rpbis value. Based on the rpbis value in Table 7 there are three items with the lowest value of discrimination power. Most of the question is a multiple-choice question. Meanwhile, essay questions have a high discrimination power value. Essay questions require students to provide discussion in answering the questions given.

The difficulty level of the questions is the opportunity to correctly answer specific ability questions, usually expressed by an index (Arifin, 2015). The test results show that there are eight items in the "easy" criteria, 11 items in the "medium" criteria, and one item in the "difficult" criteria.

The easiest questions on this test instrument are questions that test students' ability to test students' competence in explaining natural phenomena scientifically. This aligns with research conducted by (Permatasari & Fitriza, 2019), where students have the highest achievement in the aspect of competence to explain natural phenomena scientifically. The most difficult questions are questions that test students' abilities in the competence aspect of interpreting data using procedural knowledge. Based on research conducted by (Erniwati et al., 2020), students have low abilities in the aspect of competence in interpreting data scientifically. This is due to the low ability of students to process the data presented to be interpreted in other forms or to draw conclusions from the data presented. Students' ability to interpret data can be improved by using physics teaching materials that present data in various forms, such as text, tables, graphs, photos, and other forms. The learning process should ask students to compile and write experimental data and make conclusions from the data.

The study results showed that 38 participants had excellent scientific literacy skills, and 42 participants had good scientific literacy skills. This shows that more than 50% of participants understand the material of motion, force, work, and energy-related to scientific literacy. These results are in line with the results of research conducted (Dewi & Haryani, 2022), where most of the participants are in the very good and good categories, so the participant is considered to have good literacy skills.

The questions that were developed when viewed based on the competitive aspect can describe in terms of students' scientific literacy skills based on the percentage of scores obtained, as shown in the following diagram.

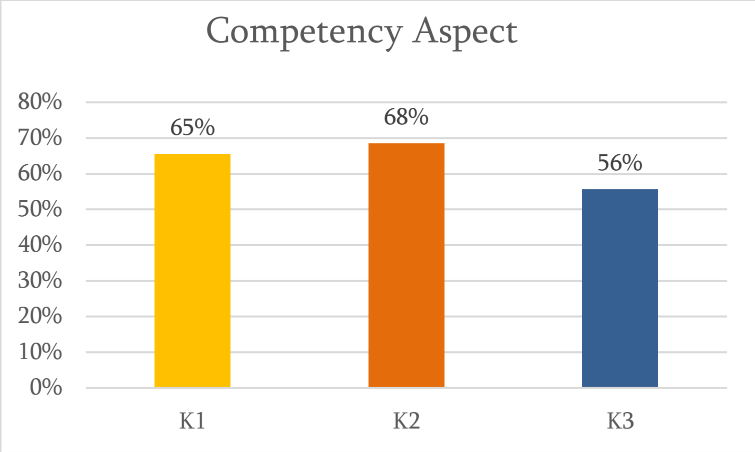

Figure 1. Student Ability on Competency Aspect

K1 : Explaining natural phenomena scientifically

K2 : Evaluation and design of scientific investigations

K3 : Interpreting scientific evidence

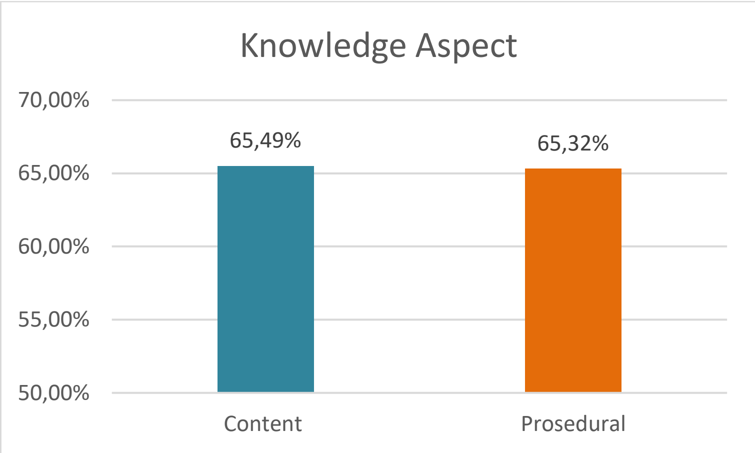

Figure 2. Student Ability on Knowledge Aspect

Based on the graph above, the conclusion is that students have very good scientific literacy skills related to knowledge relevant to real life, and students are able to explore knowledge in identifying experimental variables.

Students' scientific literacy ability can improve by increasing students' motivation in the science literacy-based physics learning process. The PISA-based scientific literacy test instrument used must consider students' abilities. The questions on the test instrument should not be too difficult but not too easy. Because questions that are too difficult for students will reduce students' motivation to work on questions. Problems that are too easy can also reduce students' motivation to work on questions because students do not feel challenged in working on questions (Arifin, 2015)

CONCLUSION AND SUGGESTION

Conclusion

The PISA (Program of International Student Assessment)-based Physics test instrument developed in this study followed the standard qualification test with a validation value of 1. The empirical validation results showed that the 20 items developed were valid with an average correlation coefficient greater than 0.40. The reliability test showed that the PISA-based physics test instrument was reliable, with a Cronbach's Alpha value of 0.816. The discrimination power is in the good and very good categories. The distribution of the difficulty level of the questions is 5% difficult questions, 55% moderate questions, and 40% easy questions.

Suggestion

1. The development of test instruments should not only focus on the material of motion, force, work, and energy so that literacy skills can be measured more optimally.
2. The test should be carried out in more than one place with more diverse characteristics of the respondents so that the respondents are more representative of the literacy abilities of high school students.

REFERENCES

Arifin, Z. (2015). Evaluasi Pembelajaran. Bandung: pt. Remaia Rosdaka.

Arikunto, S. (2012). Evaluasi Program Pendidikan. Jakarta: Bumi Aksara.

Basuki. (2014). Asesmen Pembelajaran. Bandung: PT. Remaja Rosdakarya Offset.

Dewi, A. M., & Haryani, S. (2022). Development of a Three-Tier Test Instrument Based on Scientific Literacy in Electrolyte and Non-Electrolyte Solution Topics. 5(1), 1–16.

Erniwati, E., Istijarah, I., Tahang, L., Hunaidah, H., Mongkito, V. H. R., & Fayanto, S. (2020). Kemampuan Literasi Sains Siswa Sma Di Kota Kendari: Deskripsi & Analysis. Jurnal Kumparan Fisika, 3(2), 99–108. https://doi.org/10.33369/jkf.3.2.99-108

Kemendikbud, B. (2019). Pendidikan di Indonesia belajar dari hasil PISA 2018. Pusat Penilaian Pendidikan Balitbang KEMENDIKBUD.

Lawshe, C. H. (1975). A Quantitative Approach To Content Validity. Personnel Psychology. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

OECD. (2019). Programme for international student assessment (PISA) results from PISA 2018. Oecd.

Permatasari, P., & Fitriza, Z. (2019). Analisis Literasi Sains Siswa Madrasah Aliyah pada Aspek Konten, Konteks, dan Kompetensi Materi Larutan Penyangga. EduKimia. https://doi.org/10.24036/ekj.v1i1.104087

Rizkita, L., Suwono, H., & Susilo, H. (2016). Analisis Kemampuan Awal Literasi Sains Siswa Sma Kota Malang The Analysis Of Initial Ability Of Student's Scientific Literacy In High School In Malang. Prosiding Seminar Nasional II.

Sutrisna, N. (2021). Analisis Kemampuan Literasi Sains Peserta Didik SMA di Kota Sungai Penuh. Jurnal Inovasi Penelitianitian.

Thiagarajan. (1974). Thiagarajan, Sivasailam; And Others Instructional Development for Training Teachers of Exceptional Children: A Sourcebook. Indiana Univ., Bloomington. Center for Innovation. Research and Development.