International Journal of Scientific Research and Management (IJSRM)
||Volume||12||Issue||06||Pages||3465-3475||2024||
Website: https://ijsrm.net ISSN (e): 2321-3418
DOI: 10.18535/ijsrm/v12i06.el04

Development of Multiple-choice Test Instruments to Improve Scientific Literacy in Madrasah Aliyah (MA)

Sudirman[1, 2], Ani Rusilowati [2], Endang Susilaningsih [2]

[1]Madrasah University State Islamic High School (MAN) 2 Indramayu, Indonesia.
[1,2]Faculty of Mathematics and Natural Sciences, Semarang State University, Indonesia.

Abstract
The ability to analyze and interpret scientific information is very important in the era of advanced technology. For Madrasah Aliyah (MA) students, scientific literacy is crucial for academic and future achievements, including understanding scientific concepts, interpreting data, making decisions, and communicating scientific ideas. Although scientific literacy is important, many MA students face challenges due to traditional teaching methods and inappropriate assessment. This research aims to develop and measure a multiplechoice test instrument for MA students' scientific literacy, empirically validated to improve science education and prepare students to face the modern world. Research and development (research and development) with 3D models (Define, Design, Development). Focus on scientific literacy questions with PISA indicators. Overall, it shows that the instruments used are very good and valid, and provide a positive picture of students' scientific literacy abilities, although there are still certain areas that need to be improved. Questions with poor power differences need to be repaired or replaced, and questions with sufficient power differences need to be revised. Low scores in explaining scientific phenomena require additional guidance. Regular monitoring and revision is important to maintain the quality of the instrument. This instrument is valid and reliable, but continuous improvement is needed.

Keywords: scientific literacy, instruments, validity, reliability, level of difficulty, discrimination power

Introduction

In an era characterized by rapid technological progress and scientific discovery, the ability to critically analyze and interpret scientific information has become invaluable. For students at Madrasah Aliyah (MA), developing scientific literacy is not only important for academic achievement, but also to prepare them to face future challenges and opportunities. Scientific literacy includes the knowledge and skills necessary to engage with scientific concepts and processes. Scientific literacy views the importance of thinking and acting skills which involve mastering thinking and using scientific ways of thinking in recognizing and responding to social issues [15]. This includes the ability to understand and apply scientific knowledge, interpret data and evidence, make decisions based on scientific reasoning, and communicate scientific ideas effectively. These abilities are important for personal decision making, participation in civic and cultural affairs, and economic productivity.

The Ministry of Education and Culture formulated that the 21st century learning paradigm emphasizes students' ability to find out from various sources, formulate problems, think analytically and work together and collaborate in solving problems (Ministry of Education and Culture Research and Development, 2013). Scientific literacy is an important competency in the 21st century that students really need to understand and face developments in the world around them. Scientific literacy has been widely developed in the world of education by countries such as America, Taiwan, China, Hong Kong, Australia, Germany and Chile, even developing countries such as Nigeria [13]. The Organization for Economic Co-operation and Development (OECD) defines scientific literacy as the ability to be involved in science-related problems and scientific ideas, in order to solve problems in life, as a reflective human being (OECD, 2016). Scientific literacy is closely related to students' abilities in understanding the environment, nature and surroundings. The

sustainability of nature and its surroundings is very dependent on the treatment of humans as subjects who occupy and utilize the natural environment [5]. Science education is a vehicle for students to get to know science more contextually and implement it in everyday life [9]. Scientific literacy in Indonesia was introduced in 1993 through an invitation by UNESCO to take part in the international forum on science and technological literacy for all in Paris and the realization was that a workshop on scientific and technological literacy for all in Asia and the Pacific was held in Tokyo. Scientific literacy began to be accommodated in the 2006 curriculum (KTSP) and became more clearly visible in the 2013 curriculum through inquiry activities and scientific approaches [14]. The importance of understanding and developing scientific literacy, especially biology, and attitudes towards science are relevant aspects in society [4].

Although scientific literacy is very important, many students at Madrasah Aliyah face challenges in achieving it. Traditional approaches to science teaching often focus on memorization rather than encouraging critical thinking and problem-solving skills. Current assessment practices may not adequately measure students' understanding and application of scientific concepts. Additionally, limited access to quality science teaching materials and laboratories can hinder effective science education. The focus of learning is only on mastering the material, and the assessments applied are still not appropriate so that students are only prepared to understand the knowledge [19]. To overcome this challenge, better assessment tools are needed that can measure students' scientific literacy more accurately. Effective instruments should assess multiple cognitive levels, from basic knowledge to complex analytical skills, as well as provide insight into students' strengths and weaknesses. This will support teachers in adapting teaching to meet students' needs. Current test instruments only focus on content, not on scientific literacy, such as the application of science in everyday life, critical thinking to solve problems, and the ability to carry out scientific processes [17]. PISA states that scientific literacy is "the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the changes made to it through human activity." From this explanation, scientific literacy is defined as the ability to using scientific knowledge, to identify questions and draw conclusions based on evidence to understand and help make decisions about changes that occur through human activities" [8]. Students' scientific abilities are different, so it is necessary to develop tools based on scientific literacy that can differentiate students with high abilities from students with low abilities.

This research aims to develop a multiplechoice test instrument specifically designed to improve and measure students' scientific literacy at Madrasah Aliyah. Teachers recognize scientific literacy test instruments as very important for measuring and improving students' scientific literacy but are hampered by the lack of ability to develop these instruments [3]. The instrument design developed is a multiplechoice test that corresponds to the key components of scientific literacy, validating the test instrument through empirical testing with students, and analyzing the effectiveness of the test in identifying students' scientific literacy levels. By developing a robust assessment tool, this research aims to contribute to improving science education in Madrasah Aliyah. Better assessment methods can lead to more effective teaching strategies, ultimately fostering a generation of students who are better prepared to understand and engage with the scientific aspects of the modern world.

Research Methods

Types of research

The type of research used is research and development. The development model used in this research is 4-D developed by Thiagarajan. The 4-D model consists of four stages of research and development, namely the definition stage, design stage, development stage and disseminate stage [22], with the development stage simplified to 3D (Define, Design, Development). This research was carried out in April-May 2024, at the research site at Madrasah Aliyah Negeri (MAN) 2 Indramayu, West Java. The research sample was 100 students of class The question indicators developed are in accordance with the PISA scientific literacy indicators, namely concluding and conveying findings, identifying and understanding concepts, arguing scientifically, using scientific evidence, explaining scientific phenomena, and designing and evaluating scientific investigations.

Research design

The research design aims to prepare scientific literacy assessment instruments. The research design steps include: compiling the question grid, question indicators, question form, question preparation and scientific literacy instrument design.

Data Collection Technique

The data collection techniques in this research are:

1) Conduct interviews with educators to obtain information about the problems faced during the teaching and learning process.
2) Submit the validation sheet to the expert team validator to validate the scientific literacy question instrument.
3) Conduct literature studies to support arguments and data related to the development of scientific literacy instruments.
4) Carry out measurements using tests to collect information about the characteristics of the questions to evaluate validity, reliability, level of difficulty and differentiability.

Instrument Development

The focus of this research is developing a scientific literacy multiple choice test instrument. The instrument development stages consist of (1) the definition and design stage, which includes interview guides and literacy studies; and (2) the development stage, which includes instruments built based on scientific literacy indicators which are tested with logical validation and empirical validation which includes item validation tests, reliability tests, difficulty level tests, and differential power tests as well as looking at students' scientific literacy profiles.

Data Analysis Technique

Face Validity Analysis

Face validity analysis was carried out by a validator team of experts, to analyze content, construct and language validity. This analysis uses a Likert scale validation questionnaire with 4 scales to reduce doubtful answers. The validation questionnaire uses the following formula [16].

$$PN = \frac{Validator\ Total\ Score}{Highest\ Total\ Score} \times 100\%$$

Information:

PN = Percentage Number

Actual scale = total validator score that has been averaged

Ideal scale = highest total scale score

Table 1. Validity Test Criteria

Interval	Category
3.10 – 4.00	Very suitable
2.10 – 3.00	Appropriate
1.10 – 2.00	Not suitable
0.00 – 1.00	Very inappropriade

Analysis of Students' Science Literacy Profiles

The student's scientific literacy profile is the average of the student's overall score which is measured after the questions are validated by a team of experts. The average value is measured using the formula [1]:

$$\text{Literacy Level} = \frac{Total\ Score}{Highest\ Total\ Score)} \times 100$$

The categories of students' scientific literacy profiles are explained in Table 2 [1].

Table 2. Student Science Literacy Profile Categories

Literacy level	Description
80 – 100	Very high
66 – 79	High
56 – 65	Medium
40 – 55	Low
30 – 39	Very low
< 30	Very low

Empirical Analysis

Validity test

Validity test using the product moment correlation formula with rough numbers [2] is as follows:

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right)\left(n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)}}$$

Information:

r_{xy} : correlation coefficient between variable X and variable Y, two variables that are correlated

n : number of respondents

x : average score of x

y: average score of y

The test is said to be valid if the results match the criteria. The criterion in question is parallelism, namely by using the product moment correlation technique proposed by Pearson [2].

Reliability Test

The aim of the reliability test is to see whether the questionnaire has consistency if measurements are carried out using the questionnaire repeatedly. The basis for taking the Cronbach alpha reliability test according to [21], a questionnaire is said to be reliable if the Cronbach alpha value is more than 0.6. A test produces a test that can still be said to be highly reliable. The definition of reliability is interconnected with the certainty or constancy of results [1].

Level of difficulty

Good questions are questions that are neither too easy nor too difficult. A question difficulty index of 0.0 is a difficult question, whereas an index of 1.0 indicates that the question is too easy. [2] states that to determine the difficulty index you can use the following formula:

P = B/JS

P = level of difficulty

B = the number of students who answered correctly

JS = total number of test participants

The results of the feasibility calculation are categorized according to table 3. according to Arikunto (2018).

Table 3. Difficulty Index Classification

level of difficulty	Interpretation
0.00 – 0.30	Difficult
0.31 – 0.70	Medium

0.71 – 1.00	Easy

Different Power

Differentiation power is the level of difficulty of the questions to differentiate between students who understand quickly and students who understand slowly [20]. The discrimination index measured was between 0.00 and 1.0. However, there were negative results, indicating that there were questions that were not appropriate. The discrimination index according to [1] can be measured using the following formula.

$$D = \frac{Ba}{Ja} - \frac{Bb}{Jb}$$

Information:
D = Different power
Ja = Number of participants in the upper group
Jb = Number of lower group participants
Ba = Number of participants in the upper group who answered the question correctly
Bb = Number of lower group participants who answered the question correctly
The categories of different power classification according to [1] are explained in table 4.

Table 4. Different Power Calcification

Differential Power Classification	Description
0.00 – 0.20	Poor
0.21 – 0.40	Enough
0.41 – 0.70	Good
0.71 – 1.0	Very good
D = Negative	Everything is not good

Results And Discussion

Validation of the Scientific Literacy Assessment Instrument

The validation results of expert team 1, expert team 2 and expert team 3 are explained in Table 5 below.

Table 5. Results of Instrument Validation by the Expert Team.

No	Aspect	Average score			Average score
		Validator 1	Validator 2	Validator 3	
1	Contents	3.75	3.80	3.65	3.73
2	Construct	3.74	3.74	3.57	3.68
3	Languages	3.75	3.70	3.70	3.72
Mean score		**3.74**	**3.74**	**3.64**	**3.71**
Criteria		**Very suitable**	**Very suitable**	**Very suitable**	**Very suitable**

Based on the table above, it can be explained that the average score of 3.73 indicates that the content aspect assessed is of very good quality. The validator agrees that the content is in accordance with the expected standards. The average score of 3.68 indicates that the structure and preparation of the instrument is very good and in accordance with the desired criteria, although there is slight variation in the assessment. The average score of 3.72 shows that the use of language in the instrument is very appropriate, easy to understand, and effective in conveying information. This assessment shows consistency between validators,

where scores range from 3.64 to 3.74, with the conclusion that overal the instrument is considered very suitable. All aspects were rated as “highly appropriate” by all validators, meaning the validated instrument met or even exceeded the quality expectations set by the expert team.

The results of this validation show that the instrument assessed has very good quality in terms of Content, Construct and Language. Assessments from validators consistently show that the instrument is in accordance with the expected standards. The overall mean score of 3.71 confirms that the instrument is very suitable for use and meets the specified criteria. [6] suggests that validity is a multifaceted concept and includes various types of validity, including content validity, criterion validity, and construct validity. They emphasize the importance of validity in ensuring that tests provide useful and reliable information.

Science Literacy Profile Analysis

The level of scientific literacy of Class XI Science Students at Madrasah Aliyah Negeri (MAN) 2 Indramayu shows mixed results. The results of the scientific literacy level show an average student score of 69.85. meaning that overall it is in the high category. Students' scientific literacy categories based on individual test results can be seen in Figure 1 graph below.

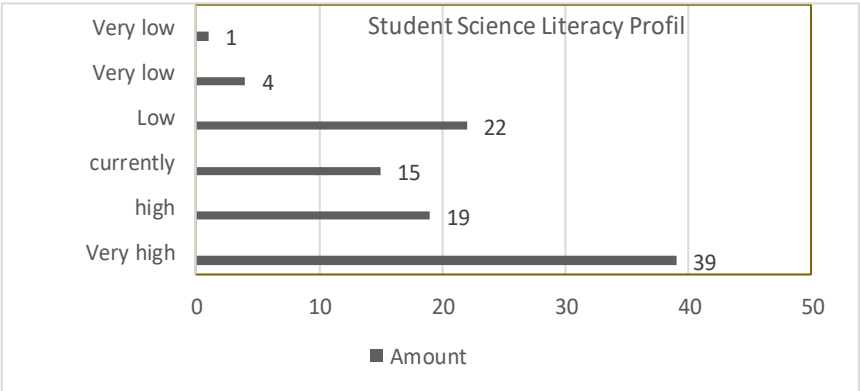

Figure 1. Student Science Literacy Profile

Based on the graph, the literacy level profile of students is as follows: 1 student is very low, 22 students are low, 15 students are medium, 19 students are high, and 39 students are very high. The majority of students have a very high level of literacy. A total of 39 students were included in the very high category, this shows that the majority of students have very good literacy skills. A small number of students have a low literacy level. A total of 22 students were in the low category, and only 1 student was in the "very low" category. This shows that although there are some students with low literacy skills, the number is relatively small. The distribution of moderate to high literacy levels, 15 students are in the "medium" category and 19 students are in the high category, shows that apart from the majority of students who have very high literacy, there are also a number of students who are at the medium to high level. Overall, this data shows that the education program at the madrasa has succeeded in improving the literacy skills of the majority of students, but there are still challenges to improving literacy among students at low levels. Based on various results but tend to be high as shown by Class XI Science students at Madrasah Aliyah Negeri (MAN) 2 Indramayu. These results cannot be separated from learning occurring more effectively and participating in a meaningful context. Scientific literacy is improved through relevant and contextual learning experiences, such as laboratory experiments, field projects, and other practical applications, which allow them to learn in real and relevant contexts [10].

The distribution of students' scientific literacy ability profiles based on the main indicators can be depicted in the graph below.

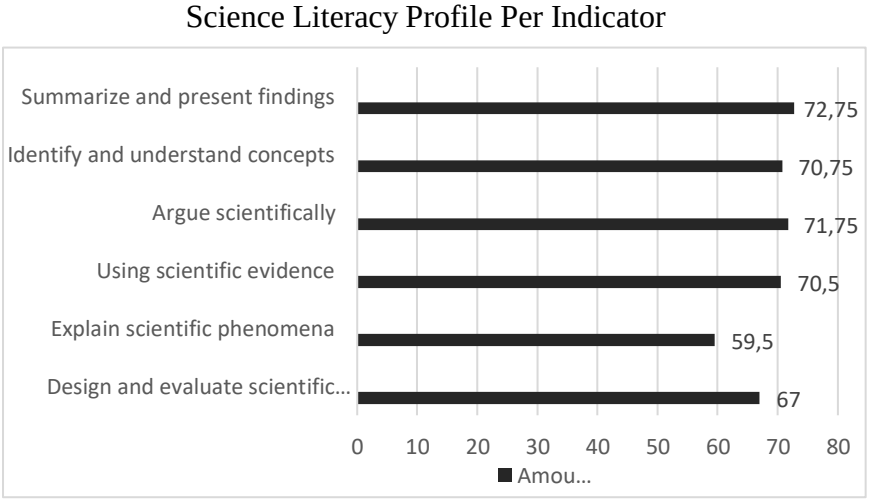

Figure 2. Science Literacy Profile Per Indicator

The graph above depicts students' scientific literacy levels based on several main indicators. Each indicator shows the percentage of understanding and skills that students have in various aspects of scientific literacy, from designing and evaluating scientific investigations to concluding and presenting findings. This data provides a comprehensive picture of students' scientific literacy abilities in various critical dimensions of science.

Students have a fairly good ability in designing and evaluating scientific investigations, with an average score of 67.0. This shows that they can plan and evaluate scientific experiments, although there is still room for improvement. An average score of 59.5 indicates that students have a moderate understanding in explaining scientific phenomena. This is the area with the lowest score, indicating a need for improvement in understanding and explaining various scientific phenomena. With an average score of 70.5, students demonstrated good ability to use scientific evidence to support their arguments or findings. This demonstrates a strong understanding of using data and scientific evidence. An average score of 71.75 indicates that students are quite skilled at arguing scientifically, constructing logical and evidence-based arguments. Students show good abilities in identifying and understanding scientific concepts with an average score of 70.75. This reflects a solid understanding of basic science concepts. The highest score was 72.75, students were very good at concluding and conveying their findings. This demonstrates a strong ability to summarize research results and present them clearly.

Overall, this graph shows that students have good scientific literacy skills, especially in concluding and conveying findings, as well as arguing scientifically, however, explaining scientific phenomena is an area that requires more attention to be improved. Explaining scientific phenomena is the area with the lowest score on the graph, indicating that students have difficulty understanding and explaining various scientific phenomena. They may need more help and guidance to improve their abilities in this aspect. According to [24] the concept of the "Zone of Proximal Development (ZPD)" indicates that students learn most effectively when they are given tasks that are beyond their current abilities but can be completed with help from others who are more skilled, such as teachers or peers. This supports the need for additional guidance in understanding complex scientific phenomena.

Empirical Analysis

Empirical analysis is a further test to determine whether the questions are suitable for use or still require revision. Empirical analysis consists of validity tests, reliability tests, difficulty level tests, and different power tests.

Validation Test

Validation analysis is used to determine whether the test instrument is valid or not. The results of the validity test can be seen in table 6 below.

Table 6. Validity Results of Question Items

Validity Index	Questions	Amount	Percentage
> 0.349 Valid	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20	20	100%
< 0.349 Invalid	-		

The validity index shows how well each item measures what it is supposed to measure. Questions with a validity index of more than 0.349 are considered valid, while questions with a validity index of less than 0.349 are considered invalid. A valid question means it meets the expected standards and measures exactly what it should measure. All questions in this test have been tested for validity and declared valid. This means that all the questions in the test are considered good and according to standards, able to accurately measure the expected aspects or competencies. Because all the questions are valid, the test as a whole is more reliable. Test results can be considered accurate and representative of the abilities or knowledge being tested. This question instrument can be used with more confidence for various purposes, such as academic evaluation or selection, because the questions are proven to be valid. Question validity is also related to the appropriate level of difficulty, not too easy or too difficult, so as to be able to differentiate between individuals who understand the material well and those who do not.

Overall, the statement that all questions are declared valid after being tested shows that the test is of good quality and can be relied upon to measure what it is supposed to measure. [11] suggests that validity is evidence that supports the interpretation of test scores for the desired purpose. They emphasize that the validation process involves collecting various types of evidence to ensure that the test actually measures the intended construct.

Reliability Test

The results of the question reliability test show that the question instrument is reliable with a Cronbach alpha value of 0.727, more than 0.6. Question reliability refers to the consistency and stability of the results provided by the question on various testing occasions. If a question or test is reliable, then the results will be consistent every time it is tested under similar conditions. Reliable questions produce consistent scores when tested on the same group under the same conditions. This means that outcomes do not change much over time if participants' abilities remain the same. Reliability shows that the measurement of the questions is stable and not affected by random factors or disturbances such as the environment, physical condition or participants' emotions.

Reliable questions can be relied upon to provide a proper evaluation, which is important in education, selection, or research, because important decisions are often based on test results. A high level of reliability indicates that the test has low measurement error and the results more closely reflect actual ability or knowledge. Reliable questions enable the reproduction of results in a variety of different contexts or situations. If questions are tested on different populations but have similar characteristics, the results will remain consistent. Test users, such as teachers, researchers, or selection administrators, can have higher confidence in reliable test results, because they know that these results can be relied on to make the right decisions.

Overall, reliability is a key aspect in the development and use of questions or tests, as it provides the basis for consistent and reliable results, which in turn supports more informed and fair decisions. Cronbach emphasized that reliability is an important prerequisite for validity. Without reliability, test results cannot be considered valid. He also developed Cronbach's alpha coefficient as a method for measuring the internal reliability of a test, which evaluates the consistency between items in the test [7]. A good instrument will provide the same measurement results and have consistent answers [18]. In the same way, a test is said to be reliable if it gives constant or consistent results if tested many times [25].

Test Difficulty Level

The level of difficulty is used to determine the level of difficulty of the test instrument. The results of the difficulty level test can be seen in the table below.

Table 7. Difficulty Level Test Results

Difficulty Level	Questions	Amount	Percentage
Difficult	-	-	-
Medium	2, 3, 5, 6, 10, 13, 17, 18, 19, 20	10	50%
Easy	1, 4, 7, 8, 9, 11, 12, 14, 15, 16	10	50%

Based on the table above, there are no questions that are included in the difficult category, so all the questions in the test are relatively easy or medium. Medium category questions are numbers 2, 3, 5, 6, 10, 13, 17, 18, 19, and 20, with a total of 10 questions, or 50% of all questions. This shows that half of the questions in the test are of medium difficulty. The easy category questions are numbers 1, 4, 7, 8, 9, 11, 12, 14, 15, and 16, also totaling 10 questions, or 50%. This means that half of the questions in the test are easy for participants to answer. From the results of this analysis, it can be concluded that the test has a balanced distribution of questions between easy questions and moderate questions. There are no questions that are classified as difficult, so it can be indicated that this test is relatively easy overall. The same proportion of medium and easy questions (50% each) shows that this test instrument is designed to measure participants' abilities with a moderate and even level of difficulty. The assessment of the level of difficulty depends on the criteria used, which can differ depending on the context and purpose of the test. [23] emphasized the importance of accurately measuring what a test aims to do and using evaluation tools that are appropriate to the measurement objectives.

Difference Power Test

Differential power is the ability of a question to measure the difference in ability between students with high abilities and students with low abilities[1]. The following are the results of the different power test of the test instrument, shown in the table below.

Table 8. Different Power Test Results

Different Power	Questions	Amount
Ugly	14	1
Just	1, 2, 8, 10, 11,12,13, 15, 16,17,18,19,20	13
Good	3, 4, 5, 6, 7, 9,	6
Very well	-	

Based on the table above, it can be explained that only one question is in the bad category, namely question number 14. This question is less effective in differentiating between participants who have high and low abilities. Maybe this question is too easy or too difficult, so almost all participants answer it the same way. Questions Numbers 1, 2, 8, 10, 11, 12, 13, 15, 16, 17, 18, 19, and 20 (13 questions) have quite different strengths. These questions have sufficient ability to differentiate between high and low ability participants. These questions can be retained on the test, but may need slight revision to improve their discrimination. Questions Numbers 3, 4, 5, 6, 7, and 9 (6 questions) are in the good differential power category. These questions are effective in distinguishing participants with high and low ability. These questions should be retained in the test without major changes.

The majority of questions were in the adequate (13 questions) and good (6 questions) categories, while only one question was in the poor category. Overall, this test has questions that vary in terms of differential power. However, there is room for improvement, especially to reduce questions with poor differential power and increase the number of questions with good and excellent differential power. The focus of improvement can be directed at questions with poor and sufficient differential power, as well as efforts to add questions with excellent differential power to increase the effectiveness of the test in measuring differences in participants' abilities. Dr. Hamzah Upu, in his book entitled "Measurement in the Education Sector" (2019), states that the distinguishing power of a question in a test or exam is an important indicator in evaluating the effectiveness of a measurement instrument. According to him, questions that have good differentiation are able to classify students based on their level of ability accurately, so that test results can provide a more valid picture of student achievement. Meanwhile, Prof. Sutama Soekarno, an expert in the field of educational evaluation, highlighted the importance of differential power analysis as a first step in improving evaluation instruments. He emphasized that good differential power test results can help teachers or test makers to compose questions that vary in level of difficulty, so that they can measure students' abilities more accurately and objectively.

Recommendation

Questions with poor differential power need to be repaired or replaced, and questions with sufficient differential power can be revised to increase the differential power. Areas with low scores, particularly in explaining scientific phenomena, require more attention with additional guidance for students to improve their abilities in this area. Regular monitoring and revision of test instruments is important to ensure their quality and effectiveness in measuring students' abilities. With results showing high validity and reliability, as well as a fairly good distribution of difficulty and differentiating power, this instrument can be relied on to evaluate students' scientific literacy. However, continuous improvement is still necessary to achieve higher quality.

References

1. Arikunto;, S. (2015). Research procedures a practical approach (Jakarta). Rineka Cipta.
2. Arikunto, S. (2018). Basics of Educational Evaluation (R. Damayanti (ed.). Jakarta: PT. Bumi Aksara.
3. Barus, R.A., Rusilowat, A., & Ridlo, S. (2024). Analysis of Needs for Development of TIMSS-Oriented Science Literacy Assessment Test Instruments for Class V Elementary School Students. JP2SD (Journal of Elementary School Thought and Development), 12 (1), 68–85.
4. Bórquez-Sánchez, E. (2024). Literasi sains dalam biologi dan sikap terhadap sains dalam sistempendidikan Chili. Penelitian dalam Pendidikan Sains & Teknologi, 1–25. https://doi.org/10.1080/02635143.2024.2320104
5. Chasanah, N., Widodo, W., & Suprapto, N. (2022). Development of a Science Literacy Assessment Instrument to Describe Student Profiles. PENDIPA Journal of Science Education, 6(2), 474–483. https://doi.org/10.33369/pendipa.6.2.474-483
6. Cohen, R. J. & Swerdlik, M. E. (2005). Psychological Testing and Assessment: An Introduction to Tests and Measurement ((6th ed.)). McGraw-Hill.
7. Cronbach, L. J. (1970). Essentials of Psychological Testing. New York: Harper & Row.
8. Harlen, W., & Qualter, A. (2004). The teaching of science in primary schools (4th ed). David Fulton.
9. Kristyowati, R., & Purwanto, A. (2019). Learning Scientific Literacy Through Using the Environment. Scholaria: Journal of Education and Culture, 9(2), 183–191. https://doi.org/10.24246/j.js.2019.v9.i2.p183-191
10. Lave, J., & Wenger, E. (1991). Situated Learning: Legitimate Peripheral Participation (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511815355
11. Linn, R.I. & Gronlund, N.E. (2000). Measurement and Assessment in Teaching. Prentice Hall.
12. OECD. (2016). PISA 2015 Results (Volume I): Excellence and Equity in Education. OECD. https://doi.org/10.1787/9789264266490-en

13. Ojimba, D.P. (2013). Scientific and Technological Literacy In Africa: Issues, Problems And Prospects' Dimensions (IPP). Educational Research International, 2(1).
14. Pertiwi, U.D., Atanti, R.D, & Ismawati, R. (2018). The Importance of Scientific Literacy In 21st Century Smp School Science Learning. Indonesian Journal of Natural Science Education (IJNSE), 01 (01), 24–29.
15. Pratiwi, S.N., Cari, C.., & Aminah, N.S. (2019). 21st Century Science Learning with Students' Scientific Literacy. Journal of Physics Materials and Learning (JMPF), 9 (1).
16. Purwanto. (2009). Evaluation of learning outcomes. Student Library.
17. Ridwan et al. (2013). Development of a Student Scientific Literacy Test Instrument on the Topic of Diversity of Living Creatures. Journal of Biology Education and Learning, 4(1), 71–78.
18. Septiani, D., Rizky, F. Y., Latifah, N., Suhani, S., & Hayashi, T. E. (2020, January 13). Analysis of Reading Comprehension Ability in Explanatory Texts in Class XI Science 5 Students of SMA Negeri 5 Karawang 2019/2020 Academic Year.
19. Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012). Discipline-based education research: Understanding and improving learning in undergraduate science and engineering. The National academies press.
20. Sugiyono. (2015). Educational research methods include quantitative, qualitative and R&D approaches. (Cet. 21). Alphabet.
21. Sujarweni, V., W. (2014). Research methodology: Complete, practical, and easy to understand. New press library.
22. Thiagarajan, S. (1974). Instructional Development for Training Teachers of Exceptional Children: A Sourcebook. https://api.semanticscholar.org/CorpusID:148925881
23. Thurstone, L.L. (1929). Theory of attitude measurement: Vol. 36(3). Psychological Review.
24. Vygotsky, L. S. (1978). Mind in Society: The Development of Higher Psychological Processes (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.; Revised ed. edition). Harvard University Press.
25. Widoyoko, E.K. (2012). Techniques for preparing research instruments. Student Son.

Author Profile

Sudirman Biology Teacher at State Islamic High School 2 Indramayu, West Java Province, Indonesia. Obtained a Bachelor of Physics Education (S.Pd) at the Indonesian Education University, Bandung and a Master of Biology Education (M.Pd.) at the University of Kuningan, West Java in the field of Electrical Engineering in 2010 and 2023 respectively. Currently pursuing a doctoral program at the State University of Semarang, Department of Science Education.