

# Development of Assessment Instrument of Scientific Literacy Ability for Students at Musamus University

Syamsul Bahri<sup>1\*</sup>, Merta Simbolon<sup>2</sup>, Kameliana Alhad<sup>3</sup>

<sup>1,2,3</sup> Department of Physics Education Universitas Musamus, Merauke, Indonesia

\*Corresponding author. Email: [syamsul\\_fkip@unmus.ac.id](mailto:syamsul_fkip@unmus.ac.id)

## ABSTRACT

This research is a type of research and development of Wilson, Oriondo, and Antonio model modification test which aims to explain the profile of the development of assessment instrument of scientific literacy ability for students at Musamus University Merauke majoring in Physics Education and Mathematics Education. The subjects in this study were 35 students majoring in Physics Education and Mathematics Education for the academic year 2019/2020 even semester. Data collection techniques in the form of tests and research instruments used are content validation questionnaire sheets, test instruments, and readability questionnaire sheets. The trials were conducted using the help of Google form. Content validation was analysed using Aiken's V formula and empirical validation was analysed using Quest assistance. Based on the results of the analysis, 46 items were declared valid in content with the condition that the value of V was 0.5 and 44 items were declared valid empirically because they showed an average value of INFIT MNSQ 1.00 SD 0.5. The reliability value shows the result of 0.67 and the average percentage of readability is 70%, which means that the reliability and readability of the instrument are categorized as high.

**Keywords:** *assessment instrument, scientific literacy*

## 1. INTRODUCTION

Assessment is a process in decision-making based on measurement results and established criteria where the two are interrelated (Purwanto, 2013). Assessment can be interpreted as one of the pedagogic competencies that must be possessed by an educator, this is according to the opinion of [1] who argues that the importance of pedagogic competencies that should be possessed by an educator, one of which is about ability in assessment.

Scientific literacy is a capacity used in scientific science, it can identify questions and draw conclusions based on the facts that have been collected and can make decisions. A person who can be said to be a scientific literate is someone who uses scientific concepts, process skills, and values in making decisions when dealing with other people or their environment can also understand the interrelation between science, technology, and society, including in economic and social development [2].

One of the assessment programs carried out by the OECD in assessing scientific literacy is through the PISA survey. The PISA survey (Program for International Student Assessment) which contains the most important abilities being tested is scientific literacy. The

presentation of the 2018 PISA data results for Indonesia in scientific ability obtained a score of 396 with a ranking of 70 out of 78 countries [3]. PISA data shows that Indonesia is among the lowest-ranked among several countries, especially in science education. This needs to be considered again considering that one of the main goals of education is the need for science education to study science and technology which is increasingly developing.

Relevant research conducted that scientific literacy-based evaluation tools still need to be developed with other themes adjusting scientific literacy-based question indicators when making questions, where the research adapts aspects of scientific literacy proposed by Chiapetta et al., 1991 [4]. Research on the assessment of scientific literacy was also carried out by Indrawati, in his journal revealed the need to develop scientific literacy assessment instruments on other materials as an assessment instrument in learning to improve students' literacy skills [5].

According to Salamah et al., that the assessment instrument is declared feasible if it meets the content validity and empirical validity requirements [4]. The feasibility of the instrument is obtained based on the

validator to determine the validation of the content, while the empirical validity can be analysed by the results of the suitability of the items that have been tested on the respondent. This is supported by a statement from Indrawati that a good assessment instrument needs to have several characteristics of item analysis which include: validity, reliability, difficulty level, distinguishing power, and problem-solving power [5].

Based on the description that has been explained, it can be concluded that the importance of developing an assessment instrument to measure students' scientific literacy skills. The development of scientific literacy assessment instruments that have been carried out by other researchers, one of which refers to Chiapetta. In this study, researchers will develop assessment instruments related to aspects of scientific literacy, as stated by Fives et al. in which these aspects consist of the role of science, scientific thinking and doing, science and society, and mathematics and science [6]. In this research, the development of the instrument contains four aspects that can be studied more deeply related to aspects of scientific literacy and developed into several indicators of scientific literacy-based questions. Researchers will conduct research entitled Development of Science Literacy Ability Assessment Instruments for Students at Musamus Merauke University.

**2. RESEARCH METHOD**

This research is a development model (Research and Development), namely a modification of the test development model from Wilson, Oriundo, and Antonio. The development steps are in the form of a test using a modified Wilson model, the Oriundo model, and the Antonio model, namely: test design, test trial, and test assembly. The data analysis technique can be seen in the feasibility of the item if it meets the eligibility criteria, namely by analysing the item validity test both in content and empirically. Content validity was analysed using Aiken's V formula and empirical validity was analysed using Quest assistance. Based on the results of Quest's analysis, it can be seen that the empirical feasibility of the assessment instrument, the difficulty level of the items, the distinguishing power of the questions, the tricking power of the questions, the reliability of the instruments. This study also analysed the readability of the scientific literacy assessment instrument using a Likert scale.

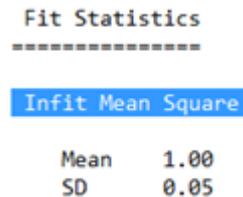
**3. RESULT AND DISCUSSION**

The data obtained from this study is based on content validation and empirical validation. The scientific literacy assessment instrument was tried out on Musamus University students majoring in Physics Education and Mathematics Education with a total of 35 students using the help of google form.

**Table 1.** The Validation Results Of The Science Literacy Ability Assessment Instrument

Validity	Item Problem	Percentage
Valid	46	96%
Invalid	2	4%

Based on the results of the research, the scientific literacy assessment instrument developed can be declared feasible or contentedly valid. The scientific literacy assessment instrument before going through the content validation stage consists of 48 items that have been validated by three validators. The data collected from the three validators used a questionnaire sheet then analyzed by calculating the Aikens'V formula. Two items have been declared invalid because both have an index of V value less than 0.5, namely item number 10 with an index value of V of 0.37 and number 27 indicating an index of V value of 0. Two invalid items are declared invalid and no repairs were made because there were still items that represented the indicators developed. So, the scientific literacy assessment instrument developed showed the results of the content validity analysis, namely that there were two invalid questions so that the questions failed and the remaining 46 items were to be tested on respondents or continued at the empirical validation stage.



**Figure. 1** INFIT MNSQ average

In the empirical validation stage, the scientific literacy assessment instrument was tested on 35 students from the Department of Physics Education and Mathematics Education at Musamus Merauke University. The trial phase carried out on students was by working on a few items and filling out a questionnaire on the readability of the instrument using the help of google form. After the test data has been collected, empirical validity analysis is carried out using Quest software. The results of the analysis show that the scientific literacy assessment instrument has a mean value of INFIT MNSQ 1.0 and SD 0.05, which means that the items as a whole are valid according to the Rasch Model.

**Table 2.** Data analysis results of problem item difficulty level

Number of Item	Difficulty Level	Criteria
4	$1 \leq x < 2$	Hard
33	$-1 \leq x < 1$	Moderate
6	$-2 \leq x < -1$	Easy
1	$x > -2$	Very Easy

The results of the analysis of the difficulty level of the items showed that 4 items were declared difficult, 33 items were declared in the moderate category, 6 items were declared easy, and 1 item was declared very easy. The items that are declared difficult on average are in the second aspect of scientific literacy, namely thinking and behaving scientifically in numbers 8, 11, 19, and 20. Meanwhile, the items stated are moderate to easy on average are in the third aspect of scientific literacy, namely science, and society. By the opinion Arikunto, it can be stated that the preparation of a good assessment instrument, is better if it needs to be dominated by items in moderate criteria such as the instrument for assessing the ability of scientific literacy that has been developed showing that there are 33 items in moderate criteria or 70% of the number of items overall [7]. Thus, based on the level of difficulty, the assessment instrument developed can be categorized as good.

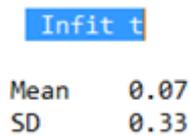


Figure 2. INFIT t average

Table 3. Results of the analysis of discriminatory power for question items

Number of Item	Discriminatory Power	Criteria
4	$x > 0,40$	Very Good
4	$0,30 \leq x \leq 0,39$	Good
11	$0,20 \leq x \leq 0,29$	Enough
25	$x \leq 0,19$	Bad

The distinguishing power of the items that have been analyzed using Quest software can be seen in the pt-biserial category with certain conditions, namely with a range of less than 0.19 to more than 0.40. If the index of distinguishing power is higher, the results of the analysis are better. Based on the results of the analysis of the distinguishing power of the items, it shows that there are 4 items that are categorized as very good, 4 items with good criteria, 11 items with sufficient criteria, and 25 items with bad criteria. Even though the average item has a distinguishing power index with a poor category, the provisions of whether or not the items can be passed can be adjusted to the criteria from Bond et al. which states that the items can be declared passed if they have INFIT t 2 [8]. Item The questions developed as a whole show an average value of INFIT t which is 0.07 with a standard deviation of 0.33, so that the assessment instrument consisting of 44 items can be declared passed and used.

Table 4. Results of the problem item confounding power analysis

Number of Item	Confounding Power	Criteria
42	$x \geq 5\%$	Good
2	$x < 5\%$	Not Good

The distinguishing power of the items that have been analyzed using Quest software can be seen in the pt-biserial category with certain conditions, namely with a range of less than 0.19 to more than 0.40. If the index of distinguishing power is higher, the results of the analysis are better. Based on the results of the analysis of the distinguishing power of the items, it shows that there are 4 items that are categorized as very good, 4 items with good criteria, 11 items with sufficient criteria, and 25 items with bad criteria. Even though the average item has a distinguishing power index with a poor category, the provisions of whether or not the items can be passed can be adjusted to the criteria from Bond, et al. which states that the items can be declared passed if they have INFIT t 2 [8]. Item The questions developed as a whole show an average value of INFIT t which is 0.07 with a standard deviation of 0.33, so that the assessment instrument consisting of 44 items can be declared passed and used.

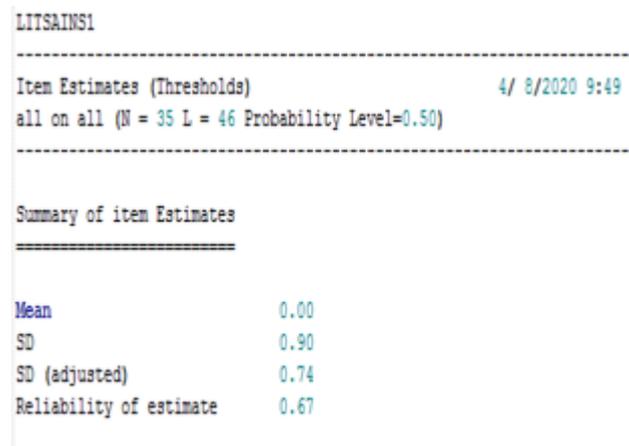


Figure 3. Results of The Reliability Analysis of The Science Literacy Assessment Instruments

The reliability index was analyzed using Quest software to determine the level of reliability coefficient in the reliability of estimate section which states that the overall assessment instrument consisting of 44 items is declared reliable according to Figure 3 which has been presented. Thus, the assessment instrument that has been developed can be stated to be able to provide consistency in measuring scientific literacy abilities. The assessment instrument that has been developed is also stated to be able to show the same results if used at different times because it has been declared reliable, according to the opinion expressed Arifin that an instrument that is declared reliable will show the same test results if tested at different [9]. Furthermore, the explanation from Arikunto states that the items can be declared reliable if the research instrument tested can provide consistent results [7].

Based on the results of data analysis obtained from data from 35 respondents who filled out the instrument readability questionnaire, the aspects that need to be addressed are the suitability of the material, aspects of

language use, and aspects of the suitability of tables and figures, so that the readability of the instrument shows better analysis results. In the assessment instrument developed, the information obtained for the legibility analysis is that the maximum percentage is 100% and the minimum percentage obtained is 37%. However, overall it shows that the results of the instrument readability analysis with a total of 35 students have an average percentage of 75% which is in the high category. In accordance with the opinion of Irmayta, it is stated that the category of percentage interpretation of the Likert scale is analyzed to determine the level of readability of the instrument [10].

#### 4. CONCLUSIONS AND SUGGESTIONS

Based on the research that has been carried out, the conclusions that can be given regarding the development of an instrument for assessing scientific literacy abilities are as follows. The instrument for assessing scientific literacy abilities that have been developed can be declared contently and empirically valid so that it is suitable to be used to measure students' scientific literacy abilities. The quality of the scientific literacy ability assessment instrument developed has been through content validation tests by three lecturers. After going through the trial phase, the scientific literacy ability assessment instrument was declared empirically valid because the mean value of INFIT MNSQ was 1.0 and SD was 0.05, which means that the overall item is valid according to the Rasch Model with the consideration that two items are invalid so that the assessment instrument remains 44 items. For the analysis of the assessment instruments that have been developed as a whole based on the difficulty level of the items in the good category, the analysis of the distinguishing power of the items is categorized as poor, the analysis of the distracting power of the items is declared good, and the reliability of the scientific literacy assessment instruments is in the high category. The results of the analysis of the readability of the scientific literacy assessment instrument with the number of respondents as many as 35 students, showed that the average percentage was 75%, which means the instrument was included in the high readability category.

#### REFERENCES

- [1] S. Bahri, R. S. Waremra, A. Reski, H. C. A. Silubun, and A. L. Rettob, "Early conditions of physics learning reviewed from teacher pedagogic competencies in SMP Negeri 11 Merauke (Ri-PNG border area)," *Int. J. Mech. Eng. Technol.*, vol. 10, no. 2, 2019.
- [2] S. Rohman, A. Rusilowati, and Sulhadi, "Analisis Pembelajaran Fisika Kelas X SMA Negeri di Kota Cirebon Berdasarkan Literasi Sains," *Phys. Commun.*, vol. 1, no. 2, p. 14, 2017.
- [3] OECD, "PISA 2018 Assessment and Analytical Framework," OECD Publishing Website, 2019. <https://doi.org/10.1787/B25efab8-En> (accessed Feb. 02, 2020).
- [4] P. N. Salamah, A. Rusilowati, and S. Sarwi, "Pengembangan Alat Evaluasi Materi Tata Surya untuk Mengukur Kemampuan Literasi Sains Siswa SMP," *UPEJ Unnes Phys. Educ. J.*, vol. 6, no. 3, pp. 7–16, 2017.
- [5] M. D. Indrawati, "Pengembangan Instrumen Penilaian Literasi Sains Fisika Peserta Didik Pada Bahasan Gelombang Bunyi di SMA Negeri 1 Gedangan Sidoarjo," *Inov. Pendidik. Fis.*, vol. 7, no. 1, 2018.
- [6] H. Fives, W. Huebner, A. S. Birnbaum, and M. Nicolich, "Developing a measure of scientific literacy for middle school students," *Sci. Educ.*, vol. 98, no. 4, pp. 549–58, 2014.
- [7] S. Arikunto, *Dasar-Dasar Evaluasi Pendidikan* (Edisi 2; R. Damayanti, Ed.). Jakarta: Bumi Aksara, 2012.
- [8] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in The Human Sciences*. In Mahwah. New Jersey: Lawrence Erlbaum Associates, 2007.
- [9] Z. Arifin, *Evaluasi Pembelajaran*. Jakarta: Rosdakarya, 2012.
- [10] E. Irmayta, R. B. Rudibyani, and T. Efkar, "Pengembangan Instrumen Asesmen Pengetahuan pada Materi Asam Basa Arrhenius," *J. Pendidik. dan Pembelajaran Kim.*, vol. 7, no. 1, pp. 63–76, 2018.