UNIVERSITAS AHMAD DAHLAN
JURNAL BIOEDUKATIKA
http://journal.uad.ac.id/index.php/BIOEDUKATIKA
2338-6630 (Print) | 2541-5646 (Online)

Development of a Scientific Literacy Test Instrument on The Immune System

Yulita Pensa [a, 1, *], Kurnia Ningsih [b, 2], Ruqiah Ganda Putri Panajaitan[c, 3]

[a,b,c] Biology Education Study Program, Fakulty Of Teacher Training And Education, Universitas Tanjung Pura Pontianak, Indonesia
[1] yulitapensa29@student.untan.ac.id *; nia_untan@yahoo.com ; ruqiah.gpp@gmail.com
* Corresponding author

ARTICLE INFO

Article history
Submission November 24, 2023
Revision February 23, 2024
Accepted June 23, 2024

Keyword:
Scientific literacy test instrument
9 TOSLS indicators
Rasch analysis

ABSTRACT

This research aims to develop a scientific literacy test instrument on immune system material. Borg & Gall research and development modified by Sugiyono is the research methodology used.. The questions developed were 25 multiple-choice questions created using nine indicators from the TOSLS instrument developed by Gormally, Brickman, and Lutz. The questions were tested on 75 class Xl students at SMA Negeri 7 Pontianak who were selected using a simple random sampling technique. Rasch model analysis was used to analyze the questions. The content validity results stated that it was valid (0.92) and had good interrater reliability (0.88). Item validity (item fit) shows 25 valid. Reliability shows the Cronbach's Alpha category is bad (0.50), the person reliability category is weak (0.46), the item reliability value is very good (0.94). The person separation value is 2 and the item separation value is 6. The difficulty of the questions (item measure) demonstrates Six questions are extremely difficult, six questions are difficult, eight questions are easy, and five questions are extremely easy. From the results of the analysis, It can be conclude that the questions are valid and reliable even though there are questions that need to be revised.

Introduction

Scientific literacy skills are considered important in 21st century education. This is inline with the viewpoint (Hidayah et al., 2017; Trilling & Fadel., 2009; Utama et al., 2019; Aiman, U., & Hasyda, S., 2020) that scientific literacy is a skill that each student needs to possess, to face the demands of 21st-century learning that are needed to keep up with the increasingly rapid development of globalization, information technology and knowledge. The skill to engage with science-related issues and with the concept of science, as a reflective citizen, is defined as scientific literacy. The term scientific literacy emphasizes the importance of science assessment on the application of scientific knowledge in the context of real-world situations (OECD.2019; Kilag, O. K., Lisao, C., Lastimoso, J., Villa, F. L., & Miñoza, C. A. 2024). Students need to have scientific literacy skills in facing the challenges of scientific development which are always related to every aspect of daily life. Students will be more accustomed to studying the problems they encounter in their lives with the theories or concepts they acquire at school (Vitasari et al., 2018; Jufrida, J., Basuki, F. R., Kurniawan, W., Pangestu, M. D., & Fitaloka, O. 2019; Aristeidou, M and Herodotou, C. 2020; Sutiani, A., Situmorang, M., & Silalahi, A. 2021).

Based on the international student assessment (PISA) program, namely a literacy study organized by the Organization for

Economic Co-Operation and Development (OECD) and the Unesco Institute for Statistics. This program is carried out to analyze the literacy skills of third grade junior high school and first grade high school students at the international stage in the aspects of reading literacy, mathematics literacy and science literacy which is held every five years. Indonesia is one of the participating countries taking part in this literacy study conducted by PISA.

The 2018 PISA results show that students in Indonesia scored below the average for OECD participants in reading, mathematics and science, Indonesia's literacy score was ranked 72nd out of 77 countries, the science literacy score was ranked 70th out of 77 countries. 78 countries. The average score of 396 is still far from the OECD average of 489, the score graph in the field of science from 2006-2018 is still relatively flat, so it must continue to be improved (OECD, 2019). In 1999 Indonesia also began participating in TIMSS which was attended by 38 countries, and in 2011 there were 79 countries as members of TIMSS, where the scientific literacy score of Indonesian students in TIMSS in 2011 was 406 (Kemendikbud, 2011: 1) and in 2015 it was in second place. 44th out of 47 countries with a score of 397, in 2019 Indonesia did not participate in this mathematics and science study (TIMSS, 2015).

Students' low level of scientific literacy can be attributed to instrument techniques that do not yet have complete criteria used to assess skills in applying scientific knowledge in real life, because types of questions that are limited to material alone will certainly not function in measuring and improving scientific literacy skills. (Mardhiyyah, 2016; Permatasari, 2022; Harrison, 2015; Diana et al., 2015)

Therefore, a scientific literacy skills instrument is needed. Although scientific literacy test instruments already exist which are standardized internationally, such as in the international research PISA and scientific literacy test instruments in the field of Biology (Gormally, et al. 2012), these instruments are included in international studies which are generally applicable.

Previous research has also been conducted on the development of a scientific literacy test instrument for science material. For example, Hasana et al. (2017) developed a test instrument for the topic of excretion and coordination systems; this test instrument can be used to measure students' literacy levels on the topic specific and improve students' literacy skills. Afterwards, research was conducted by Putri (2020) This research resulted in the development of a valid and reliable scientific literacy test instrument on the topic of the diversity of living things, followed by (Mardhiyyah, 2016) "development of a scientific literacy assessment instrument on the theme of energy", Astuti (2016) "development of an authentic assessment instrument based on scientific literacy on excretory system material", and Wulandari (2016) "analysis of scientific literacy abilities on content aspects of junior high school students on the topic of energy". This research development 25 items of scientific literacy test instrument on immune system material that has never been developed before. The scientific literacy skill indicators used in this research are the nine scientific literacy indicators contained in the TOSLS (Test of Scientific Literacy) instrument developed by Gormaly et al, (2012). The results of this research contribute to adding multiple choice questions on scientific literacy on immune system material which can be used by teachers as an assessment tool in teaching and learning activities and educational institutions can also use these questions to add types of questions with a high level of difficulty for biology competition purposes. This article will also be useful for teachers because it introduces the Rasch model using the Ministep application which makes it easy for teachers, lecturers and students majoring in education to analyze the tests they usually take. This approach is important because the analytical capabilities produced with the Rasch Model can provide richer information compared to the classical theory (score approach) that is usually used.

Scientific literacy about the immune system is required at all levels due to the rapid advancement of information and communication technology in the midst of the Covid-19 pandemic. This is necessary to learn how to distinguish fact from propaganda, analyze from fake news, and take appropriate action based on significant news reported in the media (Trilling et al, 2009; Eysenbach, G., 2020).

There is scientific literacy content in the immune system material, because the immune

system material is included in the content knowledge aspect which is assessed by PISA as one of the areas of biological science study of systems in the human body (OECD, 2018). In general, it is described in the basic competency 3.14 class.

Development can be interpreted as a scientific way of researching, designing, producing, and testing the validity of products that have been produced (Sugiyono, 2019). Scientific literacy is the skill of using scientific knowledge including nine indicators of scientific literacy according to Gormally et al. (2012), presented in Table 1.

Table 1. Indicators of scientific literacy skills (Gormally et al., 2012)

No	Scientific literacy indicators
1	Correctly identify scientific arguments
2	Conduct an effective literature search
3	Evaluating the application and misapplication of scientific information
4	Understand the elements of research design and how they impact scientific discovery
5	Create graphs that can present data
6	Read and interpret graphics from data
7	Solving problems with quantitative skills such as probability and statistics
8	Understand and be able to interpret basic statistics
9	Justify judgments and forecasts with the help of quantitative data.

Test instrument analysis can be done through Rasch modeling. According to (Sumintono & Widhiarso.,2015; Nur, L., Nurani, LA, Suryana, D., & Ahmad, A.,2020). Rasch modeling uses unique response patterns to forecast missing data. Due to this benefit, the statistical analysis of the Rasch model yields more accurate research results. It also produces standard error measurement values for the instruments that are used, which can improve computation accuracy.

That is clear from the background information provided that scientific literacy has to be raised. The primary goal of this project is to create a science literacy exam that measures students' individual suitability and skill levels.

Method

The research method used is Research and Development (R&D) using the modified Borg and Gall model (Sugiyono, 2019). This research uses seven steps out of ten research and development steps, namely potential and problems, data collection, product design, namely using test instrument preparation techniques (Mardapi, 2012), design validation, initial trial design revision and product revision. The design stage of the scientific literacy test instrument adapts the instrument development steps by (Mardapi, 2012), namely preparing test specifications and writing questions.

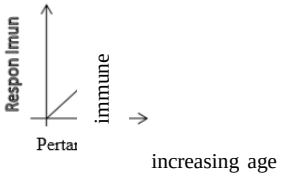
The population in this study was all students in class XI MIPA SMA Negeri 7 Pontianak for the 2020/2021 academic year, totaling 176 students. Sample selection was carried out for each class using random sampling techniques. Each class has 15 students selected by drawing lots. Data collection used interview sheet instruments, validation questionnaires, and scientific literacy tests. The test developed consisted of 25 multiple choice questions. The instrument developed refers to 9 indicators of scientific literacy according to Gormally et al. (2012) which is in table 1.

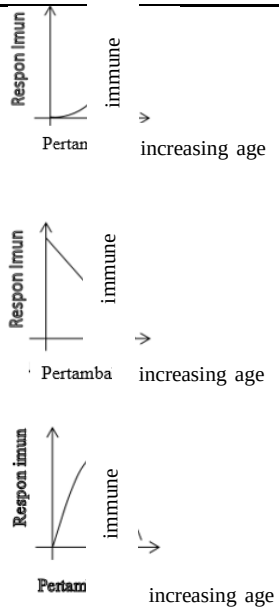
Five validators validated the content of the test before it was tested. The content validation results will be analyzed using the Aikens'V formula, and an interclass correlation coefficient (ICC) analysis will be performed using the SPSS 25 application to determine interrater reliability.. If it is declared valid and reliable, then trials can be carried out. Analysis of data from the trials used item response theory with Rasch modeling. The Rasch modeling application used is Ministep. Analysis was carried out on item validity, person reliability, item reliability, Cronbach's alpha, person separation, item separation, item difficulty level, individual ability level, and individual suitability level.

Results and Discussion

This research is development research which produces a product of a scientific literacy skills test instrument on immune system material. The test instrument is designed in a multiple choice form, totaling 31 test items with statements in the form of information at the beginning by breaking down the indicators into questions whose characteristics correspond to the details in the grid that has been created based on material rules, scientific literacy, construction and grammar. The content validity of the developed instrument is consisting of indicators of TOESL, the order (C3-C5) of the item, and the item of questions. The following are examples of questions used which are representative of cognitive levels C3-C5 which are shown in Table 2.

Table 2. Examples questions

No	Material	Objective	Cognitive Level	Gormally Et Al (2012) Ls Indicators	Question Indicators	Questions	Answer
1.	Antibody	Explain the function of antigens/antibodies for the body's defense	C3	Conduct an effective literature search	Students can evaluate the validity of journal sources about antibodies.	Background to question number 7 The benefits of antibiotics in the health sector are needed when the body's immune system really cannot fight bacteria that enter the body. A study to compare the use of antibiotics and the accuracy of administering antibiotics. The results of the study showed that when the types of antibiotics used were ampicillin, ceftriaxone, ciprofloxacin, cefepime, ceftazidime pentohydrate, cefotaxime, cefadroxil, cefditoren pivoxil, azithromycin, meropenem, gentamicin and levofloxacin, the appropriateness of antibiotics was 31 (62.0%) patients and 18 (62.0%) patients were appropriate. 36.0%) patients were inappropriate. Modification from sources: Usman, Dwi Anggara Putri., Hendra Herman., Andi Emelda. (2014). Evaluation of the use of antibiotics in community pneumonia patients at Ibnu Sina Hospital Makassar. As-Syifaa Journal Vol 06 (01) : Pg. 61-72, July 2014 ISSN : 2085-4714 (https://jurnal.farmasi.umi.ac.id/index.php/as-syifaa/article/download/34/pdf) Which of the following factors is your basis for categorizing a research article as a reliable source of knowledge? The presence of data or graphs Articles have been evaluated by experts Researcher reputation Article publisher	**B**
2.	The mechanism of formation of the immune system in the body.	Explain verbally about the mechanisms by which the immune system is formed in the body, which can be disrupted due to	C4	Create graphs that can represent data	Students can make graphs from research data about influencing factorsthe level of quality of the body's immune response	Which graph shows the relationship between increasing human age and the level of quality of the body's immune response?  	**c**

No	Material	Objective	Cognitive Level	Gormally Et Al (2012) Ls Indicators	Question Indicators	Questions	Answer
		various causes and new terms related to the immune system					
3.	Antigen and antibody mechanisms as the body's defense system	Analyze the mechanisms by which the immune system is formed, which can be disrupted due to various causes and new terms related to the immune system.	C5	Identify appropriate scientific arguments	Students can determine appropriate scientific arguments regarding the application of the working principles of antibodies and antigens in immune system disease detection tools	Two of your friends want to do a coronavirus test. First, Erwin, a child from a remote area without ODP and PDP symptoms, to complete travel requirements. The two Zakas, who had a high fever and flu, were worried about contracting the corona virus. What kind of coronavirus test would you recommend....? Erwin should undergo a rapid antigen test because it is available and the price is more affordable and the results are quite accurate and Zaka should undergo a rapid antibody test. If the result is negative, he must remain isolated and then retest. Erwin should undergo a PCR test because the results are very accurate even though they are in short supply and the price is expensive. Then Zaka should undergo a rapid antibody test. If the result is negative, he must remain isolated and then retest. Erwin should undergo a rapid antibody test because the price and availability are affordable, while Zaka should undergo a rapid PCR test because it already has the ODP and PDP criteria so it will be more accurate because the sample detected will be viral RNA and/or DNA. Erwin should undergo a rapid antigen test if the results are negative, continue with isolation and then continue with a rapid antibody test,	A

No	Material	Objective	Cognitive Level	Gormally Et Al (2012) Ls Indicators	Question Indicators	Questions	Answer
						while Zaka should undergo a rapid PCR test because he already has the ODP and PDP criteria so it will be more accurate because the sample detected will be viral RNA and/or DNA.	

Content Validation

Content validation is carried out to assess whether the test developed is valid for use or not. Content validation was carried out by 5 assessors, namely 2 lecturers from FKIP UNTAN and 3 teachers from SMA Negeri 7 Pontianak. After the validator provides an assessment, calculations are carried out using the Aiken's V formula. The results of the content validation can be seen in Figure 2.

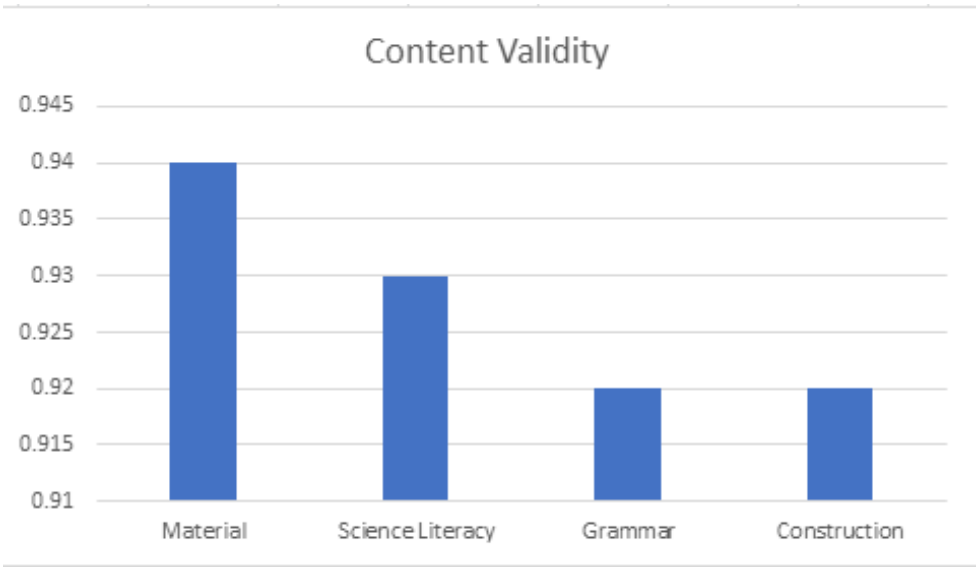


Figure 1. Content Validation Results

According to Figure 1, the average rating for the material aspect is 0.94, the scientific literacy aspect is 0.93, the construction aspect is 0.92, and the language aspect is 0.92. The average of the four aspects is 0.92. When compared with the V Aiken's table value, namely 0.87, it is categorized as valid because the calculated V Aiken > V Aiken's table. There are comments and also several suggestions on questions number 1, 3, 5, 8, 10, 11, 14, 15, 16, 21.

Revision of Scientific Literacy Test Instrument Design

The initial product of the scientific literacy test instrument was submitted to five validators. The results of the validator assessment are analyzed to produce the validity of the test content including material aspects, scientific literacy indicator content, grammar, and question construction. Of the total questions developed, namely 31 questions, only 25 (including 4 questions that have been revised, namely question items number 3, 8 , 15.18) which will be used to be tested on students in initial trials. Meanwhile, the other 6 questions, namely question items number 1, 5, 10, 11, 14, 21, could not be perfected and because of the limited capacity of the analysis program, these 6 questions were not used in the trial phase.

Inter Rater Reliability

One of the conditions for an instrument to be suitable for use is that it is reliable. After conducting content validity, inter-rater reliability was then measured. This inter rater reliability calculation uses the SPPS version 25 application based on the interclass correlation coefficient (ICC). The results of the inter rater reliability analysis can be seen in the ICC output which is shown in the average measure value obtained at 0.88. This value is then grouped based on categories referring to (Portney & Watkins, 2009), as in Table 3.

Table 3. ICC Value Category

ICC Value	Interpretation
0.00-0.50	Poor reliability
0.51-0.75	Sufficient
0.76-0.90	Good reliability
0.91-1.00	Excellent reliability

Based on Table 3, the value obtained is 0.88 so it falls into the good reliability category.

Validity of Item

Item fit describes whether or not the item functions normally to perform measurements. According to Sumintono & Widhiarso (2015), outfit mean-square, outfit z-standard, and point measure correlation values are category used to see the level of item fit. Acceptable outfit mean square (MNSQ) value: 0.5 < MNSQ < 1.5. Outfit Z-standard (ZSTD) value accepted: −2.0 < ZSTD < +2.0. Point Measure Correlation (Pt Measure Corr) value: 0.4 < Point Measure Corr < 0.85. If the question items in these three criteria are not met, it is certain that the question items are not good enough and need to be repaired or replaced. At a minimum, an item is considered fit if it meets 1 of the 3 criteria. Table 4 shows the results of the item validity analysis.

Table 4. Item Validity Analysis Results

Annotation	Question Number	Sum	Percentage (%)
Accepted	1-25	25	100
Rejected	-	0	0

Based on Table 4, it is known that of the 25 questions, 25 questions were accepted. If converted into a percentage, it is obtained that 100% of the questions are accepted or fit. According to (Palimbong et al., 2018), If the questions are declared fit, it means they meet the criteria and can make sure that the level of students' understanding is tested using appropriate and high-quality question items.

Reliability

A test is said to be reliable if the measurements are carried out repeatedly so that it will give 'steady' or 'fixed' results (Arikunto, 2012; Sürücü, L., & Maslakçi, A. 2020). In other words, reliability is the level of consistency or stability of results regarding the results of two measurements of the same thing.

According to Sumintono & Widhiarso (2015); Rahmat et al. (2020); Sudihartinih & Sufyani (2020), reliability can be determined from Cronbach's alpha value, person reliability, and item reliability. According to Sumintono and Widhiarso (2015), the person reliability value demonstrates the consistency of students' answers, the item reliability value demonstrates the quality of the question items, and the Cronbach's alpha value can measure reliability, specifically the interaction between the person and the item or the question item as a whole. The results of the reliability analysis can be seen in table 5.

Table 5. Value of Reliability

	Reliability	Category
Alpha Cronbach	0.50	Bad
Person reliability	0.46	Weak
Item reliability	0.94	Very Good

According to table 5, the Cronbach's alpha value is 0.50, indicating that the interaction between the person and the items or questions as a whole is poor. Person reliability is 0.46 and item reliability is 0.95, indicating that the consistency of student answers is poor but the quality of the questions is very good. A weak person reliability value can be caused by a lack of group variability which has implications for

the distribution of scores, namely the discriminating power index in this study, there are only two differentiating power groups, namely the high ability and low ability groups. The reliability coefficient is directly influenced by the spread of scores in the group being measured, the smaller the spread of scores, the smaller the reliability index obtained (Setiyawan, 2014; Nuryanti et al., 2018; Jumini et al., 2023; Thorndike & Robert, 2014).

Discriminating Power

Discriminating power measures how well a question can distinguish between students with high and low skill (Sumintono, B., & Widhiarso, 2015). The results of the differentiating power analysis can be seen in table 6.

Table 6. Value of Discriminating Power

	Value
Person separation	0.92
Item separation	4.03

Based on table 6, it is known that the person separation value is 0.92 and the item separation value is 4.03. Next, to see the grouping of people more carefully, an equation called strata separation is used:

$$H = \frac{[(4\text{xSEPARATION}) + 1]}{3}$$

$$H = \frac{[(4x\ 0.92\ ) + 1]}{3}$$

H=1.56

The person separation value is 0.92 so H = 1.56. The number 1.56 is rounded to 2, which means that groups of people can be divided into two groups based on the separation value. According to Sumintono & Widhiarso (2015: 12), The higher the value of person separation, the better prepared the test is because the items in it can reach students with skills ranging from high to low.

Then, the grouping of question items is calculated as follows.

$$H = \frac{[(4xSEPARATION) + 1]}{3}$$

$$H = \frac{[(4x4.03) + 1]}{3}$$

H=5. 70

The item separation value is 4.03 so H = 5.70. The number 5.70 is rounded to 6, which means that the group of questions can be divided into six groups based on the separation value. Sumintono and Widhiarso (2015) state that the greater the item separation value, the better the measurement carried out. This index is useful for defining the meaningfulness of the construct that we measure.

One of the features of Rasch modeling with the Ministep program is that it produces a map (Wright Map) which is presented in Figure 3, which depicts the distribution of student abilities and the distribution of difficulty levels of questions on the same scale. Because the logit scale on the Wright map has the same interval, this makes it easier for teachers to determine students' abilities and analyze the quality of the questions being tested, so the right information can be obtained, such as which questions students answered incorrectly a lot, so improvements can be made (Sumintono and Widhiarso, 2015). In the analysis of the Rasch model, those who have low ability are considered not to be able to work on questions with a high level of difficulty. In the Wright Map map in Figure 3, it can be seen that there are 3 questions, namely question items number 3, 5, and 15, which have a logit level. If the difficulty of a question exceeds the logit of the student's ability, the question is considered difficult and needs to be corrected.

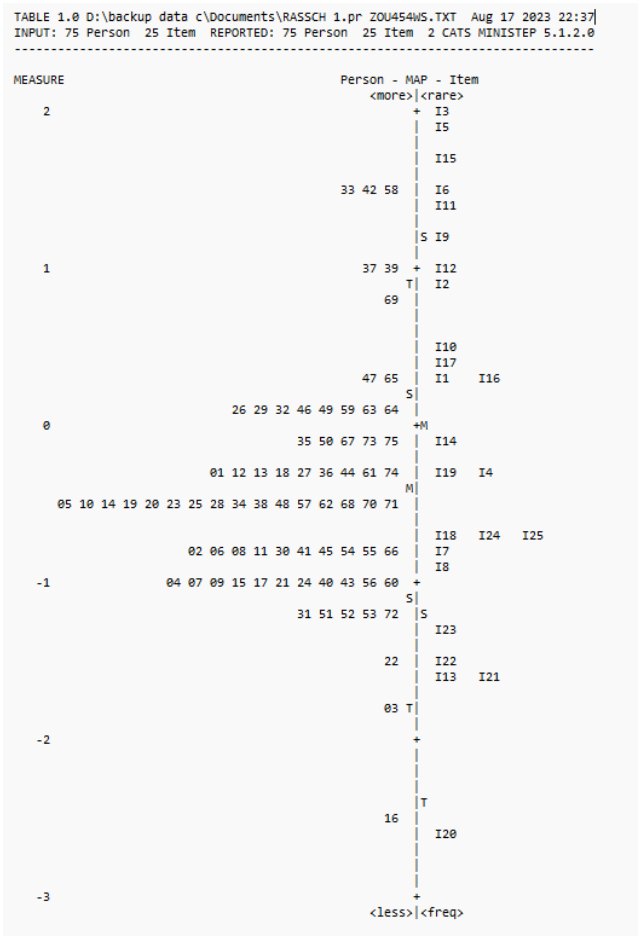


Figure 2. *Wright Map*

Difficulty Level of Question Items *(Item Measure*)

The difficulty level of a question is a value that shows how difficult and easy the question is. The results of the analysis of the level of difficulty of the questions can be seen in table 7.

Table 7. Difficulty Level Analysis Results

Category	Question Number	No. LS indicator. consecutive	Sum	Percentage (%)
Very Difficult	3, 5, 6, 9, 11, 15.	1, 2 ,3 ,4 ,6 ,8.	6	24
Difficult	1, 2, 10, 12, 16, 17.	1, 1, 5, 6, 8, 7.	6	24
Easy	4, 7, 8, 25, 14, 24,18, 19.	2, 3, 4, 9, 7, 9, 7, 5.	8	32
Very Easy	21, 20, 22, 13, 23.	1, 3, 3, 6, 9.	5	20

If the questions are linked to the scientific literacy indicators, the 6 questions in the very difficult category consist of 6 different scientific literacy indicators, namely indicator questions number 1, 2, 3, 4, 6, and 8. In the difficult question category, they consist of 2 scientific literacy indicators number 1, namely identify appropriate scientific arguments and the rest comes from 4 different scientific literacy indicators, namely scientific literacy indicators no. 5, 6, 8, and 7. The easy question category consists of 2 questions from scientific literacy indicator number 9, namely justifying conclusions, predictions, and conclusions based on quantitative data. then 2 questions from indicator number 7 that solving problems with quantitative skills such as probability and statistics, and 4 other questions come from different scientific literacy indicators, namely numbers 2, 3, 4, 5. The very easy question category consists of 2 questions with indicator number 3, evaluating the application and misapplication of scientific information and 3 questions with scientific literacy indicators 1, 6, 9. This shows that the questions developed have an inconsistent difficulty index.

The item difficulty index is obtained after testing it on students, thus allowing for differences in perception. The difficulty index for this question item really depends on the answers given by students. This is in line with Rahmat et al, (2020), that the analysis of the difficulty level of questions presents analysis results based on students' answers to the questions.

Individual Ability Level (*Person Measure)*

The level of individual ability (person measure) is an individual's ability to solve problems. This can be seen from the logit value of each individual. From this value we can compare individual ability levels with each other and sort them from highest to lowest ability level or vice versa (Sumintono & Widhiarso, 2015). The level of individual ability (person measure) is an individual's ability to solve problems. The results of the analysis of individual ability levels (person measure) can be seen in the attachment and summarized in table 8.

Table 8. Results of Test Item Ability Analysis

Category	Students	Sum	Percentage (%)
High	33L, 42P, 58P, 37L, 39P, 69P	6	8 %
Medium	47L, 65L, 26P, 29P, 32L, 46L, 49P, 59P, 63P, 64L, 35L, 50L, 67P, 73L, 75P, 01P, 12P, 13P, 18P, 27L, 36P, 44L, 61L, 74L,	24	32%
Low	05L, 10P, 14P, 19P, 20P, 23L, 25P, 28P, 34L, 38P, 48L, 57P, 62P, 68L, 70L, 71P, 02L, 06P, 08P, 11P, 30P, 41P, 45P, 54L, 55P, 66L, 04P, 07P, 09P, 15L, 17L, 21L, 24P, 40P, 43P, 56L, 60P, 31L, 51P, 52P, 53P, 72P, 22P, 03P, 16L,	45	60%

Based on Table 8, it is known that the percentage of students with high ability is 8%, medium 32%, and low 60%.

Individual Level of Conformity (*Person Fit*)

This individual's level of conformity indicates whether or not there is an out-of-the-ordinary response pattern. According to Sumintono & Widhiarso (2015), this unusual response pattern is a mismatch in the answers given based on their abilities. This can be used to determine the consistency of students' thinking and detect cheating. To see the level of suitability, criteria are used, namely the accepted Outfit mean square (MNSQ) value: 0.5 < MNSQ < 1.5, the accepted Outfit Z-standard (ZSTD) value: −2.0 < ZSTD < +2.0 and Point Measure Correlation (Pt Measure Corr) value: 0.4 < Point Measure Corr < 0.85. The results of the analysis of the level of individual suitability (person fit) can be seen in table 9.

Table 9. Results of Individual Conformity Level Analysis

	Students	Sum	Percentage (%)
Fit	42P, 39P, 69P, 66L, 15L, 47L, 58P, 70L, 33L, 03P, 12P, 17L, 43P, 63P, 20P, 26P, 10P, 60P, 14P, 21L, 05L, 68L, 19P, 23L, 65L, 22P, 54L, 18P, 28P, 67P, 07P, 30P, 01P, 74L, 29P, 32L, 06P, 73L, 04P, 41P, 56L, 24P, 31L, 75P, 09P, 48L, 27L, 40P, 52P 53P, 44L, 50L, 38P, 71P, 25P, 02L, 08P, 11P, 13P, 34L, 35L, 36P, 45P, 51P, 55P, 61L, 62P, 64L, 72P.	70	93 %
Not Fit	16L, 37L, 46L, 59P, 49P	5	7 %

From table 9, the percentage of students who are fit is 93%, while the percentage of students who are not fit or have detected unusual response patterns is 7%.

Conclusion

The scientific literacy skills test instrument that has been developed is feasible. This feasibility can be seen from the results of the content validity analysis showing that the scientific literacy skills test instrument developed is valid because V Aiken count > V Aiken table, namely 0.92 > 0.87; the interrater reliability results obtained a value of 0.881 which was categorized as good; the item validation results showed that all questions met valid criteria; The results of the reliability analysis show a person reliability value of 0.46, meaning the consistency of students' answers is weak and item reliability of 0.94, meaning the quality of the questions is very good. Cronbach's alpha value is 0.50, which means that the interaction between person and item or question items as a whole is poor; The results of the difficulty level analysis show that 6 questions are very difficult, 6 questions are difficult, 8 questions are easy, and 5 questions are very easy and the results of the discriminative power analysis show that the person separation value is 0.92, so H = 1.56. The number 1.56 is rounded to 2, which means there are two groups. person. The item separation value is 4.03 so H = 5.70. The number is rounded to 6, which means there are six groups of questions.

References

Aiman, U., & Hasyda, S. (2020). Pengaruh Model Pembelajaran Inkuiri Terbimbing Berorientasi Proses (POGIL) Berbantuan Media Realia Terhadap Peningkatan Literasi Sains dan Keterampilan Berpikir Kritis Siswa Sekolah Dasar. *Jurnal Penelitian Pendidikan Eropa* , *9* (4), 1635-1647.

Aristeidou, M and Herodotou, C. 2020. Online Citizen Science: A Systematic Review of Effects on Learning and Scientific Literacy. *Citizen Science: Theory and Practice*, 5(1): 11, pp. 7-8. DOI: https://doi.org/10.5334/cstp.224

Astuti. (2012). Pengembangan Instrumen Autentik Berbasis Literasi Sains pada Materi Sistem Eksresi. *Jurnal Unnes : Lembaran Ilmu Kependidikan*. Vol. 41 No. ISSN 0216-0847

Arikunto,S. (2012). *Dasar-dasar evaluasi pendidikan*. Jakarta: Bumi Aksara.

Diana, S., Rachmatulloh, A., Rahmawati, E. S. (2015). *Profil Kemampuan Literasi Sains Siswa SMA Berdasarkan Instrumen Scientific Literacy Assesments (SLA)*. Seminar Nasional XII Pendidikan Biologi FKIP UNS. https://media.neliti.com/media/publications/171085-ID-none.pdf

Eysenbach, G. (2020). Cara melawan infodemik: empat pilar pengelolaan infodemik. *Jurnal penelitian Internet medis* , *22* (6), e21820.

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a test of scientific literacy skills (TOSLS): Measuring undergraduates'evaluation of scientific information and arguments. *CBE-Life Sciences Education*, 11: 364-377. DOI; 10.1187/cbe.12-03-0026

Harrison, C. (2015). Assessment for learning in science classrooms. *Journal of Research in STEM Education*. *1*(2), 78-86.

Hasana, I., Saptasari, M., & Wulandari, N. (2017). Pengembangan Instrumen Penilaian Kemampuan Literasi Sains Siswa Kelas XI Materi Sistem Ekskresi dan Koordinasi di SMAN 9 Malang. *Jurnal Pendidikan Biologi*, 8(2), 52– 56.

Hidayah, R., Salimi, M., & Susiani, T. S. (2017). Critical Thinking Skill: Konsep dan Indikator penilaian. *Jurnal Taman Cendekia*, 1(2), 127–133. https://media.neliti.com/media/publications/261734-critical-thinking-skill-konsep-dan-inidi-bc4a994b.pdf

Jumini, S., Madnasri, S., Cahyono, E., Parmin, P. (2023). Analisis Kualitas Butir Soal Pengukuran Literasi Sains Melalui Teori Tes Klasik Dan Rasch Model. Prosiding Seminar Nasional Pascasarjana, 2023, 758-765. http://pps.unnes.ac.id/pps2/prodi/prosiding-pascasarjana-unnes

Jufrida, J., Basuki, F. R., Kurniawan, W., Pangestu, M. D., & Fitaloka, O. (2019). Scientific Literacy and Science Learning Achievement at Junior High School. *International Journal of Evaluation and Research in Education*, *8*(4), 630-636.

Kilag, O. K., Lisao, C., Lastimoso, J., Villa, F. L., & Miñoza, C. A. (2024). Bildung-Oriented Science Education: A Critical Review of Different Visions of Scientific Literacy. *International Journal of Higher Education Research for Innovation, Sustainability and Excellence (IJHERISE)*, *1*(1), 20-27.

Mardapi, D. (2012). *Pengukuran, Penilaian dan Evaluasi Pendidikan*. Yogyakarta: Nuha Medika.

Mardhiyyah, L., A., Ani Rusilowati, Suharto Linuwih. (2016). Pengembangan Instrumen Asesmen Literasi Sains Tema Energi. *Journal of Primary Education.* 5 (2), 150. http://journal.unnes.ac.id/sju/index.php/jpe

Nur, L., Nurani, LA, Suryana, D., & Ahmad, A. (2020). Penerapan model rasch pada instrumen pengembangan karakter siswa sekolah dasar. *Jurnal Internasional Penelitian Pembelajaran, Pengajaran dan Pendidikan* , *19* (3), 437-459.

Nuryanti,S., Masykuri, M., Susilowati, E. (2018). Analisis Iteman Dan Model Rasch Pada Pengembangan Instrumen Kemampuan Berpikir Kritis Peserta Didik Sekolah Menengah Kejuruan. *Jurnal Inovasi Pendidikan IPA*, 4 (2),224-225. http://journal.uny.ac.id/index.php/jipi

OECD (2019), PISA 2018 Assessment and Analytical Framework, PISA, *OECD Publishing*, DOI: 10.1787/b25efab8

Palimbong, J., Mujasam., & Alberto, Y. (2018). Item analysis using rasch model in semester final exam evaluation study subject in physics class x tkj SMK Negeri 2 manokwari. *Physics Education Journal*, 1(1). http://jurnal.unipa.ac.id/index.php?kpej.

Permatasari, N. (2022). Identifikasi Kompetensi Literasi Sains Peserta Didik Pada Pelajaran Ilmu Pengetahuan Alam Di Smp Negeri 43 Rejang Lebong. *Jurnal Didaktika Pendidikan Dasar*. 6 (1), 23-46, DOI: 10.26811/didaktika. v6i1.799. Doi: 10.26811/didaktika.v6i1.799

Portney, L. G., & Watkins, M. P. (2015). *Foundation of clinical research*:

applications to practice. New Jersey: Prentice Hall.

Putri, Rahmadyah Kusuma. (2020). Pengembangan Instrumen Tes Literasi Sains Siswa Pada Topik Keanekaragaman Makhluk Hidup. *Jurnal Pendidikan dan Pembelajaran Biologi* 4 (1): 71 – 78.

Robert M Thorndike. (2014). *Measurement and Evaluation in Psychology and Education ,Eighth* Edition. Edinburgh Gate Harlow: Pearson Education Limited. hal 91.

Rahmat, A. A., Hamdu, G., Nur'aeni, E., & Abdul Muiz Lidinillah, D. (2020). Pengembangan Soal Tes Tertulis Berbasis Stem Dengan Pemodelan Rasch Di Sekolah Dasar. *Metodik Didaktik*, 16(1), 29–40. https://doi.org/10.29408/didika.v6i1.2197

Setiyawan, Adhi. (2014). Faktor-Faktor yang Mempengaruhi Reliabilitas Tes. *Jurnal An Nûr*,5 (2), 351. https://jurnalannur.ac.id/index.php/An-Nur/article/view/53/55

Sudihartinih, E., & Sufyani, P. (2020). The instrument validation in plane geometry using rasch model. *Mathematics Education Journals*, 4(2). Diunduh dihttp://ejournal.umm.ac.id/index.php/MEJ.

Sugiyono. 2019. *Metode Penelitian Kuantitatif, Kualitatif, dan R & D*. Bandung: Alfabeta.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Cimahi: Trim Komunikata.

Sürücü, L., & Maslakçi, A. (2020). Validity and reliability in quantitative research. *Business & Management Studies: An International Journal*, *8*(3), 2707-2726

Sutiani, A., Situmorang, M., & Silalahi, A. (2021). Implementation of an Inquiry Learning Model with Science Literacy to Improve Student Critical Thinking Skills. *International Journal of Instruction*, 14(2), 117-138. https://doi.org/10.29333/iji.2021.1428a

TIMSS. (2015). *TIMSS Advanced 2015: Student Achievment Overview*. EIA (Energy Information Administrastion) Publishing. Diunduh di http://www.timss2015.org

Trilling, Bernie., Charles Fadel. (2009). *21st Century Skills: Learning For Life In Our Times*. United States of America. United States of America: Jossey-Bass A Wiley Imprint.

Utama, M.N., Ramadhani,R., Rohmani, S.N., & Prayitno, B.A. (2019). Profil Keterampilan Literasi Sains Siswa Di Salah Satu Sekolah Menengah Atas (SMA) Negeri Di Surakarta. *Jurnal Penelitian Pendidikan Biologi*,3 (2),57-67. http://jurnal.um-palembang.ac.id/index.php/dikbio

Vitasari, Shita Dhiyanti., dan Supahar. (2018). Assessment Instrument of Scientific Literacy Skills on Motion and Simple Machines Learning Based on Nature of Science. *International Journal of Sciences: Basic and Applied Research (IJSBAR)* 40(1),108-119. http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Wulandari, N., & Sholihin, H. (2016). Analisis Kemampuan Literasi Sains pada aspek konten Siswa SMP pada materi kalor". *Jurnal Edusains. 8(01)*.66-73. http://journal.uinjkt.ac.id/index.php/edusain