Jurnal Penelitian dan Evaluasi Pendidikan
Volume 27, No 2, December 2023 (177-188)
Online: http://journal.uny.ac.id/index.php/jpep

Validity, reliability, and item characteristics of cell material science literacy assessment instruments

Asrizal Wahdan Wilsa[1]*; Ani Rusilowati[2]; Endang Susilaningsih[3]; Jaja[4]; Veni Nurpadillah[5]
[1]STKIP Nahdlatul Ulama Indramayu, Indonesia
[23]Universitas Negeri Semarang, Indonesia
[4]Universitas Swadaya Gunung Jati Cirebon, Indonesia
[5]IAIN Syekh Nurjati Cirebon, Indonesia
*Corresponding Author. E-mail: asrizal@rocketmail.com

ARTICLE INFO

Article History
Submitted:
4 June 2023
Revised:
7 July 2023
Accepted:
18 December 2023

Keywords
instrument; item characteristics; science literacy; reliability; validity

Scan Me:

ABSTRACT

This study aims to examine the validity, reliability, and item characteristics of scientific literacy assessment instruments on cell material. The scientific literacy category encompasses the skills of discerning valid scientific arguments, assessing the credibility of sources, evaluating the proper and improper utilisation of scientific information, comprehending the components of research design and their influence on scientific conclusions, and comprehending and interpreting graphical depictions of data. The participants in the study were students enrolled in the 5th semester of the elementary school teacher education programme. The validity test involved assessing content validity with the Aiken V validity index and construct validity by Exploratory Factor Analysis (EFA). The reliability test is conducted by calculating the reliability coefficient, denoted as r, using the correlation product moment test. The characteristic test is conducted by assessing the degree of complexity and the ability to differentiate. The findings indicate that the scientific literacy assessment tool for cell material is a valid instrument, as determined through the investigation of content validity, construct validity, and item features. Therefore, the evaluation tool is appropriate for utilisation as a research instrument to gauge students' scientific literacy in cell material.

To cite this article (in APA style):
Wilsa, A. S., Rusilowati, A., Susilaningsih, E., Jaja, Nurpadillah, V. (2023). Validity, reliability, and item characteristics of cell material science literacy assessment instruments. *Jurnal Penelitian dan Evaluasi Pendidikan, 27*(2), 177-188. doi:https://doi.org/10.21831/pep.v27i2.61577

INTRODUCTION

The demands of the 21st century require that the education system be in accordance with the times. The development of science in the 21st century is oriented towards science and technology. Science education, often known as science, is an educational domain that aligns with contemporary knowledge and can be enhanced by employing scientific reasoning (Wieman, 2010). PISA research conducted by the OECD indicates that Indonesian students score 10th out of all nations in terms of scientific literacy. This demonstrates the insufficient proficiency of Indonesian students in solving PISA questions. PISA scientific literacy questions assess individuals' scientific knowledge and their capacity to recognise questions, elucidate scientific phenomena, and derive conclusions from the given evidence (Afriana et al., 2016; Hapsari et al., 2019; Khayati & Raharjo, 2020).

Scientific literacy, according to Holbrook & Rannikmae (2009), includes three aspects: understanding science in terms of concepts (what do people know), ethics or moral values (what do people value), and context (what can people do). Scientific literacy is a student's ability to apply science concepts to everyday life (OECD, 2019). Scientific literacy is not merely the ability to understand scientific knowledge but also the ability to understand scientific processes and apply them to deal with real conditions that occur in the environment (Rusilowati, 2018). Thus,

10.21831/pep.v27i2.61577

these conditions demand that the quality of science education in Indonesia improve because science education is responsible for achieving scientific literacy in the nation's children (Nyoman & Handayani, 2015).

International education is now focusing on the importance of mastering scientific literacy as the main goal of EFA (Education for All) education by UNESCO (Udompong & Wongwanich, 2014). It is necessary to carry out the assessment of students' scientific literacy abilities within a certain period of time. One of the surveys to measure students' scientific literacy skills is the PISA test, which is held every 3 years. The PISA scientific literacy assessment does not only assess a general context; more than that, the scientific literacy assessment assesses competency and knowledge in a more specific context. That is, the assessment of scientific literacy does not only assess students' understanding on the surface but also assesses students' understanding in depth (OECD, 2015). So far, evaluation tools have only emphasised content, not scientific literacy, such as applying science in everyday or contextual life, thinking to solve problems, and some scientific process abilities (Ridwan et al., 2015). Therefore, researchers should develop scientific literacy instruments to measure students' ability to apply science concepts in everyday life (Soobard & Rannikmae, 2011). Assessment is an important activity in the learning process, so it must be planned properly (Putri et al., 2022). Therefore, an instrument as a measuring tool for an assessment must be able to collect information and assess the characteristics of an object in the form of thinking abilities, attitudes, interests, and other information (Lestari & Setyarsih, 2020).

Currently, the study of scientific literacy in everyday life is very closely related to the development of technology and the technology used in science. Scientific literacy is an ability that can keep pace with the rapid development of science and technology (Mayasari et al., 2016). The scientific literacy measured in this study refers to indicators of scientific literacy according to Gormally et al. (2012). These indicators include the capacity to recognise sound scientific arguments, assess the credibility of sources, evaluate the proper and improper utilisation of scientific information, comprehend the components of research design and their influence on scientific results and conclusions, and comprehend and interpret graphical depictions of data. These indicators encompass the three indicators establishedin PISA 2018 at OECD (2019) (explaining phenomena scientifically, evaluating and designing scientific investigations, and evaluating and designing scientific investigations) but have integrated other aspects such as identifying scientific arguments, sources of information, misinformation, understanding research design, and reading data based on graphs. With the integration of several of these aspects, the level of scientific literacy that is measured is not only an aspect of knowledge but also measures the level of analysis and interpretation of students in various kinds of scientific phenomena.

The instruments developed based on these indicators refer to concepts, problems, phenomena, arguments, and data. Currently, there is a lot of science-related data that is widely published on digital platforms, but not all of the information that is disseminated is information that may not be trusted. Therefore, the instrument developed in this study also integrates students' abilities in using technology to process information, especially in scientific literacy. One example is by providing a direct source of information in the form of a website url. By tracing this information, students can directly analyse the quality of the information contained on the website and make decisions about what has been understood in that source. Students as prospective educators must possess these abilities to develop skills in identifying, analyzing, reading, interpreting, and disseminating information appropriately.

Based on the results of the preliminary study, the development of an assessment instrument in the form of questions based on scientific literacy really needs to be done. Either as an evaluation tool and reference material for teachers in developing assessment instruments or as training material in answering scientific literacy questions for students (Helendra & Sari, 2021). The development of scientific literacy test instruments can train reasoning abilities to

10.21831/pep.v27i2.61577

increase (Sinaga, 2015). Based on this background, this study aims to determine the validity and reliability of scientific literacy assessment instruments on cell material.

RESEARCH METHOD

Research on the development of scientific literacy-based evaluation instruments uses the research design of Mardapi (2008), the design is presented in figure 1 below.

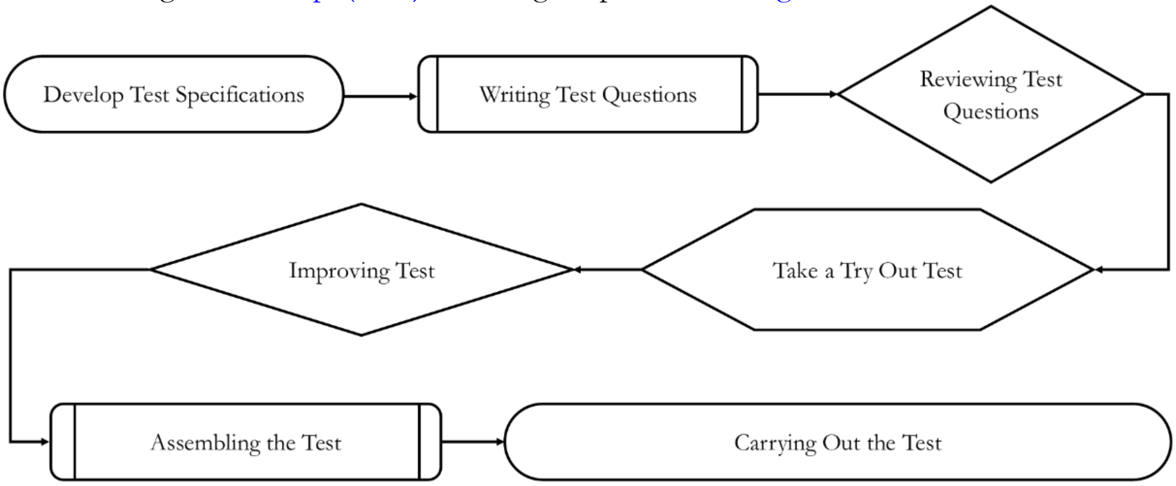

Figure 1. Stages of Developing Scientific Literacy-Based Evaluation Instruments

In detail, these steps are described as follows:

Develop test specifications

Preparation of test specifications includes determining test objectives, compiling test grids, compiling test forms, and determining test length (sentences and estimated time required).

Writing test questions

Writing questions is the act of expressing indicators by turning them into a series of questions according to the details listed in the grid that has been prepared before. Each question must be clearly formulated to find out what is being asked, and clearly explain the expected or desired answer.

Reviewing test questions

After finishing compiling the questions, the process of reviewing the questions is carried out to prevent errors that can result in difficulties for students in understanding the meaning of the questions. The process of reviewing the questions is carried out by experts (validators). The purpose of reviewing questions is to produce quality questions.

Take a try out test

Test questions aim to obtain information about the reliability, validity, level of difficulty, discriminating power, pattern of answers, and the effectiveness of the distractors on these questions. The test results will be the basis for making improvements to the questions if there is a discrepancy with the quality parameters of the questions.

Improving test

If a question does not meet the desired expectations, it can be concluded that the question is of poor quality. In this case, improvement efforts are needed in order to achieve standard questions in accordance with the guidelines set by evaluation experts.

10.21831/pep.v27i2.61577

Assembling the test

After making improvements, the next is the process of assembling questions. In this case, it is necessary to pay attention to factors that can affect the validity of the questions, such as the numbering of questions, the grouping of the questions, the layout, and other relevant matters.

Carrying out the test

After the merging or compiling process is complete and revisions have been made after the tryout, the test questions are ready to be used in administering the test to the test takers. Implementation of the test requires strict supervision to ensure that the test is carried out honestly by each participant in accordance with predetermined conditions.

The population in this study were all fifth semester students of the STKIP Nahdlatul Ulama Indramayu elementary school teacher education study program who had completed lectures on basic biology concepts. The sampling technique uses purposive sampling, namely by selecting a class with a good average value category. The number of samples in this research is 35 students. Data collection techniques used in this study were questionnaires and tests. The test instrument used was a multiples choice test which was measured using PISA development questions with indicators developed by (Gormally et al., 2012). The test questions consist of 40 test items with 5 answer choices. The indicators of scientific literacy adapted from (Gormally et al., 2012) are used as many as 5 indicators, the indicators are presented in table 1 below.

Table 1. Science Literacy Indicators

No	Indicator	Question Number	Question Item Number
1	Identify valid scientific arguments	8	1,2,3,16,17,18,31,32
2	Evaluate the validity of the source	8	4,5,6,19,20,21,37,38
3	Evaluate the use and misuse of scientific information	8	7,8,9,22,23,24,33,34
4	Understand the elements of research design and how they impact on scientific findings/conclusions	8	10,11,12,25,26,27,35,36
5	Read and interpret graphical representations of data	8	13,14,15,28,29,30,39,40

Meanwhile, the questionnaire instrument was used to determine content validity by using an instrument validation questionnaire of 40 instruments with 5 answer choices with a Likert scale. The selected validators are 6 validators who are experts in the science. Several instruments were improved based on suggestions from the validator including grammar, image quality, and effective use of sentences. The questionnaire instrument was then analyzed to determine the content validity and construct validity. Validity test includes content validity and construct validity. Content validity was tested by using the Aiken V index analysis Aiken (1985), while construct validity is Exploratory Factor Analysis (EFA) was analyzed by using IBM SPSS Statistics 27 software.
The equation for determining the Aiken V validity index is as follows:

$$V = \frac{\sum s}{n(c-1)} \tag{1}$$

Information:
V : item validity index
S : r-lo
$\sum s$: s_1+s_2+dst
N : sum of raters
c : the highest validity rating score (e.g 5)
lo : low validity rating score (e.g 1)

10.21831/pep.v27i2.61577

r : number assigned by an appraiser

Exploratory Factor Analysis (EFA) is a statistical method within the broader field of Factor Analysis (FA) that is employed to analyze the variance present among a group of observed variables or items. EFA explores the correlations among these variables, which may exhibit varying degrees of statistical significance. Consequently, this may result in a smaller number of latent factors that are not directly observed. These latent factors can be represented as a combination of the observed variables, along with some degree of error. The knowledge gained regarding the interrelationships among the observed variables serves two main purposes: reducing the number of variables or categorizing them (Iantovics et al., 2019).

The reliability test was carried out by calculating the price of the reliability coefficient using Correlation Product Moment test. The characteristic test was carried out by determining the level of difficulty, discriminating power, and the proportion or comparison of scientific literacy categories in the instrument. The developed instrument is used to measure students' scientific literacy abilities.

FINDINGS AND DISCUSSION

Content Validity

The scores obtained for the assessment instrument in the form of a Likert scale scoring rubric were analyzed using Aiken V index analysis. The results of the analysis can be categorized as valid if they meet the Aiken V index coefficient limits. The Aiken V coefficient limit requirements for 5 rating scales and 6 with an error level set at 5% obtained a minimum value based on the Aiken V validity index table of 0.79. From these data it is found that the 40 items tested can be said to be valid because they have an index that is greater than the minimum value required in the validity index table of Aiken V. In other words, the analysis results can be categorized as valid if they meet the Aiken V coefficient limit (Bashooir et al., 2018). According to Aiken V value criteria, values below 0.600 are considered to be in the lower quality category, values between 0.600 and 0.880 are classified as good, and values exceeding 0.800 are considered to be in the very good category (Suryani et al., 2017). The average score of the Aiken V index is presented in Table 2.

Table 2. Aiken V Index Average Score with 6 Validators

Item Validity Index Count	Item Validity Index V Table	Item Question	Total
0,800		3,5,6,11,13,15,16	7
0,850		1,2,4,10,12,14,17,18,19,27,28, 31,32,36,39,40	16
	0,790 (6 Validator with 5 Rating Scale)		
0,900		8,9,20,21,22,23,25,29,30,33,34, 35,37,38	14
0,950		24,26	2
1,000		7	1

Construct Validity

Factor analysis is used to determine construct validity. Factor analysis has been widely recognized in scientific circles as a quantitative social. This test is used to determine whether certain items support the factors and supporting factor variables. This test can also be used to determine which independent variables support the explanation of a certain bound. This test produces a number of factors that can explain or be an indicator of a variable. A factor occurs because the structural properties are in one relationship (Purwanto, 2004).

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v27i2.61577

The factor analysis used is Exploratory Factor Analysis (EFA) with IBM SPSS Statistic 27 software. Several conditions must be met in this analysis, namely the Kaiser Meyer Oikin Measure of Sampling Adequacy (KMO MSA) value > 0.50 and the Sig value < 0.05, the Anti-Image Correlation value > 0.50, and the Factor Loading values grouped into one group factors or components. The results of the Kaiser Meyer Olkin Measure of Sampling Adequency (KMO MSA) test are presented in Table 3 below.

Table 3. Result of Kaiser Meyer Olkin Measure of Sampling Adequency (KMO MSA) Test

Kaiser Meyer Olkin Measure of Sampling Adequency	Bartlett's Test of Sphericity		
	Approx. Chi-Square	df	Sig.
.427	24.851	10	.006

Determinant of Correlation Matrix Test. The correlation matrix is said to be interrelated when the determinant is close to 0. The calculation results show that the Determinant of Correlation Matrix is 0.006. This value is close to 0, thus the correlation matrix between the variables is interrelated. If the KMO MSA value is greater than 0.50 then the factor analysis technique can be continued. Based on the output above, it is known that the KMO MSA value is 0.427 > 0.50 and the Bartlett's Test of Sphericity (Sig.) value is 0.006 <0.05, so the factor analysis in this study can be continued because it has fulfilled the first requirement. These results are in accordance with Fadloli et al. (2023) that state a variable is said to be correlated if the MSA value is greater than 0.5.

Based on the Anti-image Matrices table, the MSA obtained from each studied (marked with a) is as follows 1) Identify valid scientific arguments 0.660, 2) Evaluate the validity of a source 0.329, 3) Evaluate the use and misuse of scientific information 0.442, 4) Understanding the elements of research design and how they impact scientific findings 0.446, and 5. Reading and interpreting graphical representations of data 0.306. Complete data is presented in Table 4.

Table 4. Anti-Image Matrices

No	Indicators	Measures of Sampling Adequacy (MSA)
1	Identify valid scientific arguments	.660[a]
2	Evaluate the validity of the source	.329[a]
3	Evaluate the use and misuse of scientific information	.442[a]
4	Understand the elements of research design and how they impact on scientific findings/conclusions	.446[a]
5	Read and interpret graphical representation of data	.306[a]

The factor analysis necessitates that the MSA values exceed 0.50. Based on the results, it is evident that the MSA value for all the variables examined is greater than 0.50, indicating that the requirements for both factor analyses have been met.

The Communalities table displays the extent to which the variable being examined may account for the factor in question. Variables are deemed to have explanatory power if their Extraction value exceeds 0.50. From the given output, it is evident that the Extraction value for all variables exceeds 0.50. Therefore, it may be inferred that all variables are capable of elucidating factors. Complete data is presented in Table 5.

Based on the Total Variance Explained output table in the "Initial Eigenvalues" section, there are 2 (two) factors that can be formed from the 5 variables analyzed. Where is the requirement to be a factor, then the Eigenvalue must be greater than 1. The Eigenvalue of Component 1 is 3.206 or > 1, so it becomes factor 1 and is able to explain 66.948% of the variation. While the Eigenvalue Component 2 is 1.298 or > 1, it becomes factor 2 and is able to

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v27i2.61577

explain 24.370% of the variation. If factor 1 and factor 2 are added up, they can explain 91.318% of the variation. The data is presented in Table 6.

Table 5. Communalities

No	Indicators	Initial	Extraction
1	Identify valid scientific arguments	1.000	.987
2	Evaluate the validity of the source		.593
3	Evaluate the use and misuse of scientific information		.999
4	Understand the elements of research design and how they impact on scientific findings/conclusions		.933
5	Read and interpret graphical representation of data		.993

Table 6. Total Variance Explained

Component	Initial Eigenvalues[a]			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	20.843	66.948	66.948	3.206	64.114	64.114
2	7.587	24.370	91.318	1.298	25.962	90.076

The next step is component matrix. This component matrix shows the correlation value or relationship between each variable and the factors that will be formed. For example: from the output above, the variable identifies a valid scientific argument, that is, the correlation value of this variable with factor 1 is 0.835, and the correlation with factor 2 is -0.537. For other variables, the way of interpreting them is the same as for variables identifying valid scientific arguments. Complete data is presented in Table 7.

Table 7. Component Matrix and Rotation Matrix

No	Indicators	Componen Matrix[a]		Rotated Component Matrix[a]	
		Component		Component	
		1	2	1	2
1	Identify valid scientific arguments	.835	-.537	**.993**	-.020
2	Evaluate the validity of the source	.755	-.149	**.721**	.269
3	Evaluate the use and misuse of scientific information	.890	.455	.519	**.854**
4	Understand the elements of research design and how they impact on scientific findings/conclusions	.929	-.264	**.929**	.262
5	Read and interpret graphical representation of data	.532	.843	.011	**.997**

By looking at the discussion above, the conclusions we can draw in this factor analysis are factor 1 with a loading value of 0.993, 0.721, and 0.929 and named the ability to identify arguments, evaluate sources, and understand research design and its impact on scientific discovery, and factor 2 with a loading value of 0.854, 0.997 and named the ability to evaluate misconceptions of information and interpret graphics data. Meanwhile, based on the Component Transformation Matrix, it shows that component 1 has a correlation value of 0.840 > 0.5, and component 2 has a correlation value of 0.840 > 0.5. Because the correlation values of all components > 0.5, it can be concluded that the two factors formed are feasible to summarize the five variables analyzed. The data is presented in Table 8.

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v27i2.61577

Table 8. Component Transformation Matrix

Component	1	2
1	.840	.542
2	-.542	.840
Information :		
Extraction Method : Principal Component Analysis		
Rotation Method : Varimax with Kaiser Normalization		

Reliability

There are three methods for calculating test reliability, namely: First, the Equivalent Method, second is the Test-Retest Method, and third is the Split-half method. In the ANATES application, to calculate the reliability of the test using the Split-half Method, namely dividing the odd-even and dividing the initial and final items. The equation for calculating the reliability of tests on anates applications according to Wiguna (2021), namely:

$$r_{11} = \frac{rx^{1/2\ 1/2}}{r +^{1/2\ 1/2}} \quad (2)$$

Information :
r_{11} : test reliability coefficient
$r^{1/21/2}$: correlation coefficient of odd-even scores (XY correlation)

Based on the reliability test of 40 test items from 35 samples using the ANATES application, the average result was 18.00, standard deviation was 8.81, XY correlation was 0.86, and test reliability was 0.92. These results are then compared with the test instrument reliability interpretation criteria, namely:
1. If $r_{count} > r_{table}$ the data is said to be reliable
2. If $r_{count} > r_{table}$ the data is said to be unreliable

The minimum reliability value based on r table for 40 multiple choice test items with an error rate of 0.5% is 0.312. The calculated r value obtained based on analysis is 0.92 > 0.312, so it can be said that the scientific literacy assessment instrument on cell material is a reliable instrument. That is, a test can be said to have a high level of confidence if the test can give consistent results. It means the extent to which a test can be trusted to produce a score that is consistent or does not change (Arikunto, 2008).

Item Characteristics

Characteristics of the items include the level of difficulty of the items that are good with the composition of easy, medium and difficult questions spread proportionally according to the subject matter being tested and the different power index of the items that are good are able to distinguish between groups of students with high abilities and groups of students with low abilities , so that the results of the evaluation of student learning will describe the actual learning outcomes of students. In this study, the test of the characteristics of the items used was the Classical Test Theory (CTT) using the ANATES software. Classical test theory emphasizes the raw score of a single exam being produced. The raw score indicates a person's ability. From this raw score, various analyzes and interpretations can be produced according to the needs of the study being conducted (Sumintono & Widiharso, 2015).

Based on the test of the characteristics of the items from the 35 samples tested, it was found that the differentiating power of the questions with the score category according to

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v27i2.61577

Arikunto (2008) was negative (not good), 0.00-0.20 (bad), 0.21-0.40 (sufficient), 0.41-0.70 (good), and 0.71-100 (very good). Based on this analysis, 14 items (35%) with very good discriminating power were obtained, 15 (37%) were good, 6 (15%) were sufficient, 2 (5%) were bad, and 3 (8%) were not good. Based on these results it can be said that the instrument items have good discriminating power because as many as 35 questions fall into the very good, good, and sufficient categories, this is in accordance with (Arifin, 2012) that the higher percentage value of the discriminating power coefficient, the better the item to be able to differentiate student abilities. The complete data is presented in Table 9.

Table 9. Results of the Analysis of the Discriminating Power of the Questions

No	Category	Question Item Number	Quantity	Percentage
1	Very good (0,71-100)	1,2,3,4,5,7,8,12,14,17,20,22,23,25	14	35%
2	Good (0,41-0,70)	6,9,10,13,15,21,26,27,28,29,31,32,33,35,36	15	37%
3	Sufficient (0,21-0,40)	11,16,18,19,30,40	6	15%
4	Bad (0,00-0,21)	24,37	2	5%
5	Negative (Not good)	34,38,39	3	8%

Meanwhile, the level of difficulty of the questions based on analysis with the ANATES software from 35 samples and 40 questions showed that the questions tested were included in the questions with medium and difficult categories. The number of questions in the difficult category is 6 questions (15%) and the questions in the medium category are 34 questions (85%). Based on these results, most of the questions are in the medium category. Thus, it can be said that the questions tested are questions that have a good level of difficulty. This is in accordance with Arikunto (2012) which states that the items used tend to use questions that are not too easy and not too difficult (moderate). Elviana (2020) also states that if the number of items made is more in the difficult category, it will result in students not having the interest and motivation to try again in giving answers because the questions are beyond their ability. Complete data is presented in Table 10.

Table 10 . The results of the analysis of the difficulty level of the test questions

No	Category	Question Item Number	Quantity	Percentage
1	Hard	4,7,16,24,34,38	6	15%
2	Medium	1,2,3,5,6,8,9,10,11,12,13,14,15, 17,18,19,20,21,22,23,25,26,27,28, 29,30,31,32,33,35,36,37,39,40	34	85%

Scientific literacy skills are a very important aspect for students to master, this is related to the way they understand the environment, health, economics and other problems of modern society which depend on technology and the progress and development of science (Wulandari, 2016). Teachers as learning designers have the competence to develop assessment or evaluation instruments for certain skills, one of which is students' scientific literacy skills (Putri et al., 2020). Assessment or evaluation is a tool used to measure the achievement of learning objectives (Fraenkel et al., 2012). Preparing an evaluation instrument based on scientific literacy is an effort to measure students' literacy abilities, especially in the field of science or natural science education (Fu'adah et al., 2017). Apart from that, the development of scientific literacy assessment instruments is carried out so that students can be trained in writing questions or solving problems in life by applying science (Martinah et al., 2022).

CONCLUSION

Based on the tests of content validity, construct validity, reliability, and item characteristics of the cell material scientific literacy assessment instrument, it can be concluded that the 40 test items included in the valid criteria. Furthermore, construct validity using the EFA analysis test

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v27i2.61577

obtained 2 factors, factor 1 named the ability to identify arguments, evaluate sources, and understand research design and its impact on scientific discovery, and factor 2 named the ability to evaluate misconceptions of information and interpret graphics data. Based on the component transformation matrix of the two factors formed, it can be concluded that it is feasible to summarize the five variables analyzed. The assessment instrument being analyzed is a reliable instrument with a reliability index of 0.92. Meanwhile, the level of item difficulty was included in the moderate category, discriminatory power was included in the good and very good categories, and the correlation of item scores with the total score was mostly included in the very significant category.

REFERENCES

Afriana, J., Permanasari, A., & Fitriani, A. (2016). Project Based Learning Integrated to STEM to Enhance Elementary School's Students Scientific Literacy. *Jurnal Pendidikan IPA Indonesia*, *5*(2), 261–267. https://doi.org/10.15294/JPII.V5I2.5493

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Arifin, Z. (2012). *Evaluasi Pembelajaran*. PT. Remaja Rosdakarya.

Arikunto, S. (2008). *Prosedur Penelitian Suatu Pendekatan Praktek*. Rineka Cipta.

Arikunto, S. (2012). *Prosedur Penelitian*. Rineka Cipta.

Bashooir, K., Bashooir, K., & Supahar, S. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(2), 219–230. https://doi.org/10.21831/pep.v22i2.19590

Elviana, E. (2020). Analisis Butir Soal Evaluasi Pembelajaran PAI Menggunakan Program Anates. *Jurnal MUDARRISUNA: Media Kajian Pendidikan Agama Islam*, *10*(2), 209–224. https://doi.org/10.22373/JM.V10I2.7839

Fadloli, Pamungkas, R., Sumarjoko, T., & Marlina, R. (2023). Analisis Faktor Pengembangan Instrumen Pengukuran Keterampilan Abad 21 Siswa Sekolah Menengah Atas. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, *6*(3), 1995–2003. https://doi.org/10.54371/JIIP.V6I3.1802

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to DesIgn and Evaluate Research in EducatIon (8th ed.)*. Mc Graw HIll.

Fu'adah, H., Rusilowati, A. (2017). Pengembangan Alat Evaluasi Literasi Sains untuk Mengukur Kemampuan Literasi Sains Siswa Bertema Perpindahan Kalor dalam Kehidupan. *Lembaran Ilmu Kependidikan*, *46*(2), 51–59. https://doi.org/10.15294/lik.v46i2.11350

Gormally, C., Brickman, P., & Lut, M. (2012). Developing a test of scientific literacy skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *CBE Life Sciences Education*, *11*(4), 364–377. https://doi.org/ 10.1187/cbe.12-03-0026

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

Hapsari, E. E., Sumantri, M. S., & Astra, I. M. (2019). Strategi Guru Meningkatkan Hasil Belajar Menggunakan Pendekatan Saintifik di Sekolah Dasar. *Jurnal Basicedu*, *3*(3), 850–860. https://doi.org/ 10.31004/basicedu.v4i1.171

Helendra, H., & Sari, D. R. (2021). Pengembangan Instrumen Asesmen Berbasis Literasi Sains Tentang Materi Sistem Ekskresi dan Sistem Pernapasan. *Jurnal Ilmiah Pendidikan Profesi Guru*, *4*(1), 17–25. https://doi.org/10.23887/jippg.v4i1.29860

Holbrook, J., & Rannikmae, M. (2009). The Meaning of Scientific Literacy. *International Journal of Environmental and Science Education*, *4*(3), 275–288. http://www.ijese.com/

Iantovics, L. B., Rotar, C., & Morar, F. (2019). Survey on establishing the optimal number of factors in exploratory factor analysis applied to data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(2), e1294. https://doi.org/10.1002/widm.1294

Khayati, D. N., & Raharjo. (2020). Pengembangan Instrumen Tes Berbasis Literasi Sains untuk Memetakan Critical Thinking dan Practical Skills Siswa pada Materi Sistem Peredaran Darah Kelas XI SMA. *Berkala Ilmiah Pendidikan Biologi (BioEdu)*, *9*(3), 433–442. https://doi.org/10.26740/bioedu.v9n3.p433-442

Kusuma Putri, R., Pendidikan, Y., & Medan, S. A. (2020). Pengembangan Instrumen Tes Literasi Sains Siswa Pada Topik Keanekaragaman Makhluk Hidup. *Diklabio: Jurnal Pendidikan Dan Pembelajaran Biologi*, *4*(1), 71–78. https://doi.org/10.33369/diklabio.4.1.71-78

Lestari, D., & Setyarsih, W. (2020). Kelayakan Instrumen Penilaian Formatif Berbasis Literasi Sains Peserta Didik pada Materi Pemanasan Global. *IPF: Inovasi Pendidikan Fisika*, *9*(3), 561–570. https://doi.org/10.26740/ipf.v9n3.p561-570

Mardapi, D. (2008). *Teknik Penyusunan Instrumen Tes dan Non Tes*. Mitra Cendikia Offset.

Martinah, A. A., Mubarok, V., Miarsyah, M., Ristanto, R. H. (2022). Pengembangan Instrumen Tes Literasi Sains Berbasis Kontekstual pada Materi Pencemaran Lingkungan. *Bioedusiana: Jurnal Pendidikan Biologi*, *6*(2), 192–218. https://doi.org/10.37058/bioed.v6i2.3251

Mayasari, T., Kadarohman, A., Rusdiana, D., & Kaniawati, I. (2016). Apakah Model Pembelajaran Problem Based Learning dan Project Based Learning Mampu Melatihkan Keterampilan Abad 21? *Jurnal Pendidikan Fisika Dan Keilmuan (JPFK)*, *2*(1), 48–55. https://doi.org/10.25273/jpfk.v2i1.24

Nyoman, N., & Handayani, L. (2015). Membangun Masyarakat Melek Sains Berkarakter Bangsa Melalui Pembelajaran. *Prosiding Seminar Nasional MIPA*. https://ejournal.undiksha.ac.id/index.php/semnasmipa/article/view/10307

OECD. (2015). *PISA 2015 Result in Focus. Paris: OECD-PISA.* https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf

OECD. (2019). *PISA 2018 Results (Volume I): What Students Know and Can Do. In OECD Publishing: Vol. III.*

Purwanto, Dr. (2004). Analisis Faktor: Konsep, Prosedur Uji dan Interpretasi. *Jurnal Teknodik*, 153–169. https://doi.org/10.32550/teknodik.v4i15.388

Putri, A. A. A., Hussain, H., & Ramdhani. (2022). Pengembangan Instrumen Tes Literasi Sains Pada Dimensi Pengetahuan Materi Asam Basa. *SCIENCE: Jurnal Inovasi Pendidikan Matematika Dan IPA*, *2*(4), 536–547. https://doi.org/10.51878/science.v2i4.1797

Ridwan, M., Mardhiyyah, L. A., & Rusilowati, A. (2015). Pengembangan Instrumen Asesmen Dengan Pendekatan Kontekstual untuk Mengukur Level Literasi Sains Siswa.

Rusilowati, A. (2018). Asesmen Literasi Sains: Analisis Karakteristik Instrumen dan Kemampuan Siswa Menggunakan Teori Tes Modern Rasch Model. *Pros Semin Nas Fis Univ Riau ke-3 2018;1-14*.

Sinaga, T. (2015). Pengembangan Soal Model Pisa Mata Pelajaran Ilmu Pengetahuan Alam Terpadu Konten Fisika untuk Mengetahui Penalaran Siswa Kelas IX. *Jurnal Inovasi dan Pembelajaran Fisika*, *2*(2), 194–196. https://doi.org/10.36706/jipf.v2i2.2623

Soobard, R., & Rannikmae, M. (2011). Assessing Student's Level of Scientific Literacy Using Interdisciplinary Scenarios. *Science Education International*, *22*(2), 133–144. https://files.eric.ed.gov/fulltext/EJ941672.pdf

Sumintono, & Widiharso. (2015). *Aplikasi pemodelan RASCH pada assessment pendidikan*. Tim Komunikata Publishing Hous.

Suryani, H., Kartowagiran, B., & Jaelani, J. (2017). Development and validity of mathematical learning assessment instruments based on multiple intelligence. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *21*(1), 93–103. https://doi.org/10.21831/pep.v21i1.15286

Udompong, L., & Wongwanich, S. (2014). Diagnosis of the Scientific Literacy Characteristics of Primary Students. *Procedia - Social and Behavioral Sciences*, *116*, 5091–5096. https://doi.org/10.1016/j.sbspro.2014.01.1079

Wiguna, S. (2021). *Aplikasi Anates dalam Evaluasi Pembelajaran*. CV. Pena Persada.

Wieman, C.E. (2010). Science education for the 21st century: using the insights of science to teach/learn science. Technical Symposium on Computer Science Education. https://doi.org/10.1145/1734263.1734333

Wulandari, N. (2016). Analisis Kemampuan Literasi Sains Pada Aspek Pengetahuan dan Kompetensi Sains Siswa SMP Pada Materi Kalor. *EDUSAINS*, *8*(1), 66–73. https://doi.org/10.15408/es.v8i1.1762

ISSN (print) 2685-7111 | ISSN (online) 2338-6061