Validity and reliability scientific literacy-based physics assessment instrument on atomic nucleus and radioactivity material

Ladika Zuhrotul Wardi[1]*, Mukhayyarotin Niswati Rodliyatul Jauhariyah[2]

[1]Magister of Science Education, Faculty of Mathematics and Natural Sciences, State University of Surabaya

[2]Physics Education Study Program, Faculty of Mathematics and Natural Sciences, State University of Surabaya

*Corresponding author, email: ladika.23016@mhs.unesa.ac.id

Article History
Received: 9 July 2024
Revised: 4 November 2024
Accepted: 13 December 2024

Keywords
assessment instruments
atomic nucleus
radioactivity
scientific literacy

Abstract
Indonesia has a low level of scientific literacy, so it is necessary to make efforts for teachers to develop appropriate assessments or assessment instruments. In this research, a physics assessment instrument will be developed based on scientific literacy on atomic nuclei and radioactivity with the aim that the resulting instruments are good and feasible to use. This research is a type of development research. The feasibility analysis of the assessment instruments in this study included validity, reliability, level of difficulty, and differentiability. The assessment instrument will be validated by three expert lecturers as validators. The test subjects in this study were 35 first-year students of the Physics Education Study Program, at Surabaya State University. Based on the results of data analysis, it can be concluded that the scientific literacy-based physics assessment instrument on atomic nuclei and radioactivity developed has a logical validity value for the content domain of 100% in the very valid category, the construction domain of 92.85% in the valid category, and the language by 85.71% with very valid category. As for the empirical validity value, 50% of the questions are in the valid category and 50% are in the invalid category. The reliability value is 0.492 which means reliable. Furthermore, for the difficulty level of the items, 20% of the questions were in the easy category, 65% of the questions in the moderate category, and 15% of the questions in the difficult category. As well as for the discriminating power of the items, 10% of the items were in the good category, 35% were in the sufficient category, and 55% were in the less category.

Wardi, L. Z., & Jauhariyah, M. N. R. (2025). Validity and reliability scientific literacy-based physics assessment instrument on atomic nucleus and radioactivity material. *Momentum: Physics Education Journal*, *9*(1), 109-119. https://doi.org/10.21067/mpej.v9i1.10450

1. Introduction

One of the most important drivers accelerating growth across all domains is the 21st century. The education sector is actively working to create learning models that are more in line with the needs of the twenty-first century in this age of globalization (Azizahwati et al., 2023; Deta et al., 2023). It can be observed from the development of these situations and conditions, various efforts have been made to improve the quality and quality of education. One of the goals of learning in the 21st century is to prepare students to be able to enter the world of work in a wider scope (Rizaldi et al., 2020). The 21st century can be faced by creating quality human resources so they can compete in the international arena (Dakhi et al., 2020). In the 21st century learning process requires several skills that must be possessed by students, so that learning must be student-centered (Parcon, 2022; Tyan et al., 2020). One of the skills needed in the 21st century is scientific literacy skills (Apriyani et al., 2021). Scientific literacy is defined as the ability to act as a reflective individual, using his scientific knowledge so that he can engage with science-related problems and provide scientific ideas to solve real-life problems and make decisions related to the surrounding circumstances (Nurhayati et al., 2023). Scientific literacy can also be explained as the ability to understand scientific knowledge and its life application (Lestari et al., 2018).

The ability of students to have scientific literacy in Indonesia is in the low category and has decreased. This is based on the results of the assessment Program for International Student Assessment (PISA) from 2015 to 2022 where Indonesia is ranked lower than the participating

doi: 10.21067/mpej.v9i1.10450

ISSN: 2548-9135

countries. In 2015 students in Indonesia achieved a score of 403 out of the overall average score of participating countries, namely 493. Then in 2018, students in Indonesia only achieved a score of 396 out of the overall average score of participating countries, namely 489. This means that Indonesia experienced a significant decline between 2015 and 2018 (Sholikah & Pertiwi, 2021; Wahyu et al., 2020). While in 2022, Indonesian students reached a score of 383 from the international average score of 485 (OECD, 2023). Several factors can affect students' scientific literacy levels, including teaching materials, learning models, learning environment, student assignments, and scientific literacy-based assessment instruments (Radzi et al., 2023; Saputra et al., 2024). Assessment instruments based on scientific literacy do not only focus on real situations but also mastering concepts in life skills (Aina et al., 2020; Rosana et al., 2020).

Currently, the Indonesian education system generally uses two types of curriculum, namely the 2013 curriculum and the "Merdeka Belajar" curriculum. The 2013 curriculum clearly emphasizes student skills that can be integrated into real-life. As the new curriculum or the "Merdeka Belajar" curriculum progresses, some new programs or policies replace the National Examination (UN) program with the Minimum Competency Assessment (AKM) program. AKM is a student-oriented assessment that measures basic literacy and numeracy skills. There are only three materials tested in this AKM, namely literacy, numeracy, and also strengthening character education. Whereas literacy and numeracy materials refer to PISA questions (Shofiyah et al., 2021). This AKM has the intention to measure student competence in depth and not just content mastery. This is following the assessment of scientific literacy which aims not only to measure aspects of students' knowledge but also to emphasize students' scientific competence (Dewi et al., 2021; Ke et al., 2021).

Scientific literacy itself is closely related to physics, especially its relationship with surrounding life (Apriyani et al., 2021). Most students still think that physics only contains theories and is rarely associated with application in everyday life (Rini & Aldila, 2023; Wulandari et al., 2022). The physics cognitive abilities possessed by students are also classified in the low category (Kurniawan et al., 2020). In addition, many students also think that physics material is mainly related to atomic nuclei and radioactivity which contains analyzing the characteristics of atomic nuclei, analyzing radioactivity events, and the use and dangers of radioactivity is one of materials that is very difficult and boring (Cardoso et al., 2020). Studying physics requires the ability to assemble an understanding of concepts and also understanding their applications in solving problems in everyday life (Erfianti et al., 2019). In the atomic core and radioactivity, the material itself is closely related to life. For example, in the medical field, some radioisotopes are used to diagnose patients and treat medical devices. By utilizing gamma rays, Co-60 can be used as a more practical and effective sterilizer for medical devices (Astuty & Angkejaya, 2022).

AKM itself is a form of effort from the government to develop scientific literacy skills possessed by students. But before that, it is necessary to make efforts from the teacher through the development of an appropriate assessment or assessment instrument. The self-assessment instrument aims to measure the abilities possessed by students. Assessment instruments can be used properly if they can measure the abilities of the students to be measured (Jacob & Rothstein, 2016; Sumarni et al., 2018). The minimum criteria for an assessment instrument can be said to be good, namely valid and reliable, having varying levels of difficulty, and having differentiating power. Research on the development of science literacy assessments on atomic nuclei and radioactivity is needed because these topics are complex and abstract, demanding deep understanding from students. This assessment is not only designed to measure basic knowledge, but also critical thinking skills and application of concepts by students related to issues such as radiation safety and nuclear technology applications (Appiah-twumasi, 2024). In contrast to previous studies that generally focus on general physics topics such as force, motion, temperature, and heat (Anibo et al., 2022; Danar et al., 2025; Elisa et al., 2019). This study offers a customized assessment instrument to evaluate students' science literacy in the context of atomic nuclei and radioactivity, providing a more specific approach in measuring students' science understanding. Meanwhile, the scientific literacy competencies that will be used in the development of assessment instruments in this study refer to the 2018 scientific literacy competencies by PISA which are described in Table 1.

Table 1. PISA Science Literacy Competence 2018

No.	Science Literacy Competence
1	Explaining phenomena scientifically
2	Evaluating and designing scientific inquiry
3	Interpreting data and evidence scientifically

(OECD, 2019)

The problems to be discussed in this study are formulated as follows: (1) What is the validity and reliability of the developed science literacy-based physics assessment instrument on atomic nuclei and radioactivity material? (2) What is the level of difficulty of the developed science literacy-based physics assessment instrument on atomic nuclei and radioactivity material? (3) What is the distinguishing power of the developed science literacy-based physics assessment instrument on atomic nuclei and radioactivity material?

2. Method

This research is a type of development research. This study aims to produce a physics assessment instrument based on scientific literacy competence in atomic nuclei and radioactivity that is good and feasible to use. The assessment instrument developed was in the form of 20 descriptive questions covering three scientific literacy competencies according to PISA, namely explaining phenomena scientifically, evaluating and designing scientific investigations, and interpreting data and evidence scientifically. The feasibility analysis of the assessment instruments in this study included validity, reliability, level of difficulty, and differentiability. The test subjects in this study were 35 first-year students of the Physics Education Study Program, at Surabaya State University. The validity test conducted in this study includes logical validity and empirical validity.

The logical validity test was carried out by three Unesa Physics Department lecturers as expert lecturers who would assess it based on a Likert scale with a score of 1-4, then analyzed using the mode of each validation aspect. The results obtained will be calculated by the percentage of matches between the validators using the equation Percentage of Agreement (PoA) as follows:

$$Percentage\ of\ Agreement = 100\%\ x\ (1 - \frac{A-B}{A+B}) \quad (1)$$

with:
A = Frequency of the aspect assessed by the observer by giving a high frequency
B = The frequency of the aspect assessed by the observer by giving a low-frequency
Acceptance criteria percentage of agreement is if the results obtained are ≥ 75% (Borich, 1994)

Furthermore, empirical validity is determined using the correlation formula product moment as follows:

$$r_{xy} = \frac{N\ \Sigma XY - (\Sigma X)(\Sigma AND)}{\sqrt{\{N\Sigma X^2 - (\Sigma X)^2\}\{N\Sigma AND^2 - (\Sigma AND)^2}} \quad (2)$$

information:
r_{xy} = Correlation coefficient for each item item
N = Number of trial respondents
X = Sum of item scores
AND = Sum of total scores

The criterion for testing the validity of the items is by calculation r_{xy} then compare with rproduct moment on the table, when $r$count > $r$table then it can be said that the item is valid (Arikunto, 2012).

According to Arikunto (2012), a test has high reliability if the test gives consistent results. Essay form test reliability is measured by the coefficient formula Alpha Cronbach as follows:

$$r_{11} = \left(\frac{n}{n-1}\right)\left(1 - \frac{\Sigma\sigma i^2}{\sigma t^2}\right) \quad (3)$$

information:
r_{11} = Reliability Value
$\Sigma\sigma i^2$ = Total variance of the item scores
σt^2 = Varians total

The test reliability test criteria are from calculations $r11$ then compared with rproduct-moment on the table, it can be said reliable when $r$count > $r$table (Sugiyono, 2015).

The difficulty level of the item is a measure of the ease or difficulty of an item based on the ability of the test subject. So that the level of difficulty can be seen from the ability of the test subjects to answer the items being tested. To calculate the difficulty level of the item description questions, use the following formula:

$$P = \frac{Mean}{Shoes\ max} \tag{4}$$

information:
P = Difficulty Index
Mean = Average score for each item
Shoes Max = The maximum score that has been set for each item

The difficulty level is categorized based on the difficulty index which ranges from 0.00 to 1.00. The higher the difficulty index, the easier the item is because almost all test subjects can answer, and vice versa (Arikunto, 2012).

The discriminating power of items is the ability of an item to distinguish between the upper and lower groups or distinguish high-ability test subjects from low-ability test subjects. To calculate the discriminating power of the item description questions use the following formula:

$$D = \frac{S_A - S_B}{Score\ Max} \tag{5}$$

information:
D = Discrimination Index
SA = Upper group average
SB = Lower group average
Score Max = Maximum score

Discriminating power is categorized based on the index of discrimination. The higher the discrimination index, the more effective the items are in distinguishing high- and low-ability test subjects, and vice versa (Arikunto, 2012).

3. Results and Discussion

The development of the assessment instrument begins with analyzing the initial needs or conditions, the analysis covers the curriculum to basic competencies to determine the material used in this study. After obtaining the results of the analysis, then proceed to the stage of designing the assessment instrument. The assessment instrument was developed based on predetermined indicators and scientific literacy competencies by PISA 2018. In this case, the assessment instrument includes a question grid, a table of item specifications, and an assessment rubric. Furthermore, the assessment instrument will go through a logical validation stage and then try out to obtain empirical validity, reliability, level of difficulty, and discriminatory power. Table 2 presents an initial development containing data on the distribution of item items based on scientific literacy competencies.

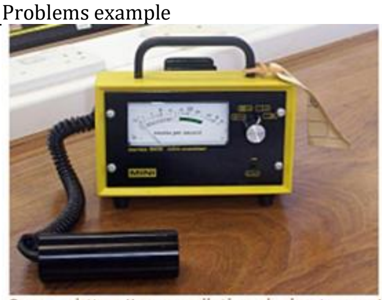
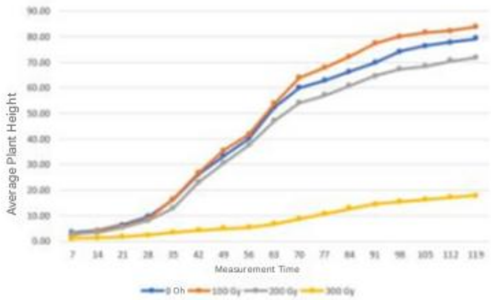
Table 2. Data on the Distribution of Question Items Based on Scientific Literacy Competencies

Items	Explain phenomena scientifically	Evaluate and design scientific investigations	Interpret data and evidence scientifically
1	√		
2			√
3			√
4			√
5			√
6			
7			
8	√	√	√
9		√	
10	√		
11			√
12	√		
13	√		√
14			
15			
16	√		
17			√
18		√	
19	√	√	
20			√
Sum	6	5	9

Based on Table 2, the initial assessment instrument consisted of 20 questions in the form of a description with a percentage distribution of 30% for the competence to explain phenomena scientifically, 25% for the competence to evaluate and design scientific investigations, and 45% for the competency to interpret data and evidence scientifically. Table 1Table 3 shows some examples of questions developed based on scientific literacy competencies:

Table 3. Examples of questions based on scientific literacy competencies

Science Literacy Competence	Problems example
Explain phenomena scientifically	 REPUBLIKA.CO.ID, MOSCOW Quoting a spokesman for the Zaporizhzhia nuclear power plant (NPP) in Ukraine, the Russian news agency RIA reported that the baseline radiation level has not changed there, still within safe levels. The largest nuclear power plant in Europe caught fire during an attack by Russian troops. Separately, RIA, citing the Ukrainian emergency service, said that the fire broke out outside the perimeter of the nuclear power plant. One of the generating blocks has been shut down. Source: https://www.republika.co.id Question: Based on the discourse above, explain what impacts will occur if humans are exposed to nuclear radiation?

Science Literacy Competence	Problems example
Evaluate and design scientific investigations	 Source: https://www.radiation-dosimetry.org/ Nuclear radiation counts in today's technology are increasingly developing. This is due to the increasing need for mapping the internal structure of objects in line with the support for increasing nuclear technology capabilities in various fields. Nuclear spectroscopy systems and radiation counters function to investigate and analyze a radiation source. By measuring the distribution of energy and various other information related to the radiation source. Question: Based on the information above, explain the purpose and data obtained from the experiment that can be carried out using the Geiger Muller detector!
Interpret data and evidence scientifically	 Source: Tia et al., 2021 The image above is a graph of the average physiological growth of chili plant height against measurement time for each dose given. Question: Based on the graph above, what information can be obtained?

Based on the table above, are examples of questions from the three indicators of scientific literacy. For the first problem with the indicator of scientifically explaining phenomena using questions with initial information related to a nuclear disaster, students are expected to be able to explain the impact resulting from the disaster, namely exposure to nuclear radiation. For the second problem with indicators evaluating and designing scientific investigations using questions with information and pictures related to nuclear radiation counter experiments or Geiger Muller detectors, students are expected to be able to design experiments that are carried out by explaining the objectives and experimental data resulting from the experiment. Then for the third problem with indicators of interpreting data and evidence scientifically using questions with information in the form of graphs that describe the physiology of the average height of chili plants irradiated against the time of measurement, students are expected to be able to interpret graphs into information or data.

The initial assessment instrument was validated by three expert lecturers as validators based on aspects of the content, construction, and language domains using a validation questionnaire. The results of the logical validity of scientific literacy-based physics assessment instruments on the atomic nucleus and radioactivity material are described in Table 4.

Based on the logical validation results in Table 4, it can be concluded that the average PoA for each domain has a percentage of ≥ 75% so the assessment instrument developed is included in the valid category. Viewed from the realm of content with a percentage of 100%, the category is very valid, the realm of construction, with a percentage of 92.85%, is a valid category, and the domain of language, with a percentage of 85.71%, is a very valid category. This shows that the developed scientific literacy-based physics assessment instrument meets valid criteria and can be used both in terms of material, construction, and language.

Table 4. The results of the logical validity of the instrument

Contents	Aspect number	Modus	Criteria
	1	4	Very Valid
	2	3	Valid
	3	3	Valid
	4	4	Very Valid
Modus			Very Valid
PoA Average			100%
Construction	Aspect number	Modus	Criteria
	5	4	Very Valid
	6	3	Valid
	7	3	Valid
	8	3	Valid
Modus			Valid
PoA Average			92,85%
Material	Aspect number	Modus	Criteria
	9	4	Very Valid
	10	4	Very Valid
	11	3	Valid
	12	4	Very Valid
Modus			Very Valid
PoA Average			85,71%

Table 5 shows some suggestions given by the validator during the assessment instrument validation process.

Table 5. Suggestions and improvements from the validator

Suggestions and improvements	Question items
The use of the preposition di- in the word "*diatas*" should be separated by a space to become "*di atas*"	2, 3, 4, 5, 10, 11, 13, 14, 15, 17, 18, 19
Some of the information presented, such as images, tables, and graphics, is blurry or unclear, so it needs to be corrected	10, 13, 18
Some of the language used is still unclear and specific, so it needs to be explained in detail so as not to cause confusion	8, 10, 15, 17

Then the assessment instrument was tested on students who would then be analyzed. The results of the empirical validity of the physics science literacy test items were analyzed using Ms. Excel is described in Table 6.

Table 6. The results of the empirical validity of the item questions

Criteria	Question Items	Amount
Valid	3, 4, 7, 8, 10, 11, 13, 15, 18, 19	10
Invalid	1, 2, 5, 6, 9, 12, 14, 16, 17, 20	10

Based on Table 6, analyzed from $r_{\text{count}} > r_{\text{table}}$ then the 20 items on the scientific literacy test that have been developed yield the results of 10 valid questions and 10 invalid items or a percentage of 50% valid and 50% invalid.

Table 7. Literacy competency data for valid questions

Science Literacy Competence	Question Items
Explain phenomena scientifically	8, 11, 19
Evaluate and design scientific investigations	7, 15, 18
Interpret data and evidence scientifically	3, 4, 10, 13

The results of the reliability of scientific literacy-based physics assessment instruments analyzed using Ms. Excel are described in Table 8.

Table 8. Instrument reliability results

R_{11}	R_{table}	Criteria
0.492	0.334	Reliable

Based on Table 8, the results of the reliability analysis of the developed assessment instrument obtained an R 11 of 0.492 which means $r_{count} > r_{table}$ so it can be concluded that the instrument is reliable. This is in accordance with the research by Zuhdi et al., (2024).Then for the processing of the difficulty level of the physics science literacy test items, the results are obtained in Table 9.

Table 9. The results of the difficulty level of the item questions

Criteria	Question Items	Amount
Easy	1, 2, 15, 19	4
Medium	3, 4, 6, 7, 8, 9, 10, 11, 13, 14, 16, 17, 20	13
Difficult	5, 12, 18	3

Based on Table 9, the 20 items are categorized into three levels of difficulty namely easy, medium, and difficult. The results of the item difficulty analysis showed a balanced distribution across the three categories of difficulty. A total of 20% of the items were in the easy category to assess students' basic understanding, ensuring a foundation of prior knowledge. Most of the items, 65%, were in the medium category and served as the main tool for testing deeper understanding and application of the concepts learned. Meanwhile, another 15% of questions were categorized as difficult, requiring in-depth analytical thinking to identify higher-ability students. It can be seen that most of the questions are categorized as medium. This is in accordance with the research by Zuhdi et al., (2024) and Arafah et al., (2021) states that a good question is a question that is not too easy and not too difficult. Meanwhile, research conducted by Suprapto et al., (2020) stated that the dominant instrument on questions with difficult categories.

For the processing of the discriminating power of the physics science literacy test items, the results are obtained in Table 10.

Table 10. The results of the discriminating power of the item questions

Criteria	Question Items	Amount
Less	1, 5, 6, 8, 9, 12, 14, 16, 17, 18, 20	11
Enough	2, 4, 7, 10, 13, 15, 19	7
Good	3, 11	2

Based on the table above, the 20 items are categorized into three categories of discriminating power, namely good, sufficient, and lacking. Judging from the differentiating power of the developed assessment instrument, 10% of the questions were in the good category, 35% were in the sufficient category, and 55% were in the poor category. Questions with good discriminating power criteria have abilities that can distinguish students who have understood competence from students who have not understood competence (Mulyana & Desnita, 2023). Questions with poor discriminating power criteria will result in the work of students in the upper group and students in the lower group getting equally high scores or equally low scores (Arafah et al., 2021; Suprapto et al., 2020). Figure 1 is a documentation of the assessment instrument trials conducted on students.

Figure 1. Documentation of testing the assessment instrument

4. Conclusion

The developed physics assessment instrument based on scientific literacy on the atomic nucleus and radioactivity material has a logical validity value for the content domain of 100% with the very valid category, the construction domain is 92.85% with the valid category, and the language domain is 85.71% with the category very valid. As for the empirical validity value, 50% of the questions are in the valid category and 50% are in the invalid category. The reliability value is 0.492 which means reliable. Furthermore, for the difficulty level of the items, 20% of the questions were in the easy category, 65% of the questions in the moderate category, and 15% of the questions in the difficult category. As well as for the discriminating power of the items, 10% of the items were in the good category, 35% were in the sufficient category, and 55% were in the less category. So, it can be concluded that from the scientific literacy-based physics assessment instrument on atomic nuclei and radioactivity developed, the results of instruments that are valid, reliable, and can be used to measure students' scientific literacy competency abilities are obtained. Valid and reliable instruments ensure that assessment results truly reflect students' science literacy, especially on complex and abstract material. In addition, this instrument can support a curriculum that is more in line with science literacy needs and has the potential to become a reference for national assessment standards, thus helping to prepare students for future science challenges.

Author Contributions

All authors have equal contributions to the paper. All the authors have read and approved the final manuscript.

Funding

No funding support was received.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Aina, J. K., Opeyemi, B. A., & Olu, A. M. (2020). Assessment of scientific literacy skills of college of education students in Nigeria. *American Journal of Social Sciences and Humanities*, *5*(1), 207–220.

Anibo, Z., Gebeyehu, D., & Assefa, S. (2022). Assessing college students ' factual , conceptual , and procedural knowledge of newton ' s laws of motion. *Momentum: Physics Education Journal*, *6*(2), 199–213.

Appiah-twumasi, E. (2024). Scaffolding as a cognitive load reduction strategy for teaching atomic and nuclear physics. *Momentum: Physics Education Journal*, *8*(2), 194–209. https://doi.org/10.21067/mpej.v8i2.9580

Apriyani, Y., Supriyati, Y., & Margono, G. (2021). The Influence of Learning Models on Scientific Literacy in Physics Course: A Meta-Analysis Research. *International Journal of Science and Society*, *3*(4), 44–51. https://doi.org/10.54783/ijsoc.v3i4.392

Arafah, K., Amin, B. D., Sari, S. S., & Hakim, A. (2021). The Development of Higher Order-Thinking Skills (HOTS) Instrument Assessment in Physics Study. *Journal of Physics: Conference Series*, *1899*(1), 12140. https://doi.org/10.1088/1742-6596/1899/1/012140

Arikunto, S. (2012). *Dasar-Dasar Evaluasi Pendidikan*.

Astuty, E., & Angkejaya, O. W. (2022). Pelatihan Sterilisasi Alat Dan Bahan Medis Pada Anggota Tim Bantuan Medis Vertebrae Fakultas Kedokteran Universitas Pattimura. *Society : Jurnal Pengabdian Masyarakat*, *1*(5), 284–290. https://doi.org/10.55824/jpm.v1i5.137

Azizahwati, A., Januarti, J., Sari, S. W., Sari, R. A., Ranti, L., & Septyowaty, R. (2023). Meta analysis of the effect of STEM application on higher order thinking skill in science learning. *Momentum: Physics Education Journal*, *7*(1), 154–163. https://doi.org/10.21067/mpej.v7i1.7959

Borich, G. D. (1994). *Observation Skill for Effective Teaching*. Macmillan Publishing Company.

Cardoso, P. S. S., Nunes, M. C. S., Silva, G. P. S., Braghittoni, L. S., & Trindade, N. M. (2020). Conceptions of high school students on atomic models, radiation and radioactivity. *Physics Education*, *55*(3), 35030. https://doi.org/10.1088/1361-6552/ab7fc6

Dakhi, O., Jama, J., Irfan, D., Ambiyar, & Ishak. (2020). BLENDED LEARNING: A 21ST CENTURY LEARNING MODEL AT COLLEGE. *International Journal of Multi Science*, *1*(7), 50–65.

Danar, I., Irawan, A., & Kurniawan, B. R. (2025). Effect of isomorphic problems with feedback to reduce student misconceptions on simple harmonic motion. *Momentum: Physics Education Journal*, *9*(5), 1–13. https://doi.org/10.21067/mpej.v9i1.10048

Deta, U. A., Ayun, S. K., Laila, L., Prahani, B. K., & Suprapto, N. (2023). PISA science framework 2018 vs 2025 and its impact in physics education:Literature review. *Momentum: Physics Education Journal*, *8*(1), 95–107.

Dewi, C. A., Erna, M., Martini, M., Haris, I., & Kundera, I. N. (2021). The Effect of Contextual Collaborative Learning Based Ethnoscience to Increase Student ' s Scientific Literacy Ability. *Journal of Turkish Science Education*, *18*(3), 525–541.

Elisa, N., Kusairi, S., Sulur, S., & Suryadi, A. (2019). The Effect of Assessment for Learning Integration in Scientific Approach Towards Students ' Conceptual Understanding on Work and Energy. *Momentum: Physics Education Journal*, *3*(2), 103–110.

Erfianti, L., Istiyono, E., & Kuswanto, H. (2019). Developing Lup Instrument Test to Measure Higher Order Thinking Skills (HOTS) Bloomian for Senior High School Students. *International Journal of Educational Research Review*, 320–329. https://doi.org/10.24331/ijere.573863

Jacob, B., & Rothstein, J. (2016). The measurement of student ability in modern assessment systems. *Journal of Economic Perspectives*, *30*(3), 85–107. https://doi.org/10.1257/jep.30.3.85

Ke, L., Sadler, T. D., Zangori, L., & Friedrichsen, P. J. (2021). Developing and Using Multiple Models to Promote Scientific Literacy in the Context of Socio - Scientific Issues. *Science and Education*, *30*, 589–607. https://doi.org/10.1007/s11191-021-00206-1

Kurniawan, E. S., Mundilarto, M., & Istiyono, E. (2020). Synectic HOTS oriented: Development of teaching materials for high school physics learning. *Universal Journal of Educational Research*, *8*(11), 5547–5554. https://doi.org/10.13189/ujer.2020.081158

Lestari, T. P., Sarwi, S., & Sumarti, S. S. (2018). STEM-based Project Based Learning model to increase science process and creative thinking skills of 5th grade. *Journal of Primary Education*. https://journal.unnes.ac.id/sju/index.php/jpe/article/view/21382

Mulyana, V., & Desnita. (2023). Empirical Validity and Reliability of the Scientific Literacy Assessment Instrument Based on the Tornado Physics Enrichment Book. *Jurnal Penelitian Pendidikan IPA*, *9*(5), 3961–3967. https://doi.org/10.29303/jppipa.v9i5.3290

Nurhayati, Y., Sopandi, W., & Riandi, R. (2023). Scientific Literacy Profile of Pre-Service Elementary School Teacher Students. *Jurnal Penelitian Pendidikan IPA*, *9*(7), 5474–5480. https://doi.org/10.29303/jppipa.v9i7.3510

OECD. (2019). PISA 2018 Assessment and Analytical Framework, PISA. In *OECD Publishing*. https://doi.org/10.1787/b25efab8-en

OECD. (2023). Equity in education in PISA 2022. In *PISA 2022 Results (Volume I): The State of Learning and Equity in education* (Vol. 1). https://www.oecd-ilibrary.org/education/pisa-2022-results-volume-i_03c74bdd-en

Parcon, R. E. (2022). Task-Based Instruction in teaching 21st Century literature from the Philippines and the world: Learners' views and reflections. *ELT Forum: Journal of English Language Teaching*, *11*(1), 41–51. https://doi.org/10.15294/elt.v11i1.52493

Radzi, M., Bakar, A., Halim, L., & Mohamad, N. (2023). Factor Influencing Students ' Scientific Literacy : An Exploratory Factor Analysis. *International Journal of Academic Research in Business & Social Sciences*, *13*(12), 5056–5075. https://doi.org/10.6007/IJARBSS/v13-i12/20393

Rini, E. F. S., & Aldila, F. T. (2023). Practicum Activity : Analysis of Science Process Skills and Students ' Critical Thinking Skills. *Integrated Science Education Journal*, *4*(2), 54–61. https://doi.org/10.37251/isej.v4i2.322

Rizaldi, D. R., Fatimah, Z., Sciences, N., & Study, E. (2020). The Correlation of Digital Literation and STEM Integration to Improve Indonesian Students ' Skills in 21st Century. *International Journal of Asian Education*, *01*(2), 73–80.

Rosana, D., Widodo, E., Setyaningsih, W., & Warno, D. S. (2020). DEVELOPING ASSESSMENT INSTRUMENTS OF PISA MODEL TO MEASURE STUDENTS ' PROBLEM -SOLVING SKILLS AND SCIENTIFIC LITERACY IN JUNIOR HIGH SCHOOLS. *Jurnal Pendidikan Sains Indonesia*, *8*(2), 292–305. https://doi.org/10.24815/jpsi.v8i2.17468

Saputra, W. T., Rochintaniawati, D., & Agustin, R. R. (2024). The Role of Learning Models in Enhancing Scientific Literacy : A Critical Review Evaluation. *Prisma Sains: Jurnal Pengkajian Ilmu Dan Pembelajaran Matematika Dan IPA*, *12*(3), 462–476.

Shofiyah, N., Mauliana, M. I., Istiqomah, I., & Wulandari, R. (2021). STEM Approach: The Development of Optical Instruments Module to Foster Scientific Literacy Skill. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika*, *5*(2), 92–103. https://doi.org/10.36312/esaintika.v5i2.388

Sholikah, L., & Pertiwi, F. N. (2021). Integrative Science Education and Teaching Activity Journal Analysis of Science Literacy Ability of Junior High School Students Based on Programme for International Student Assessement (PISA). *Integrative Science Education and Teaching Activity Journal*, *2*(1), 95–104.

Sugiyono, S. (2015). *Statistik Untuk Penelitian*. Alfabeta.

Sumarni, W., Supardi, K. I., & Widiarti, N. (2018). Development of assessment instruments to measure critical thinking skills. *IOP Conference Series: Materials Science and Engineering*, *349*(1). https://doi.org/10.1088/1757-899X/349/1/012066

Suprapto, E., Saryanto, Sumiharsono, R., & Ramadhan, S. (2020). The Analysis of Instrument Quality to Measure the Students ' Higher Order Thinking Skill in Physics Learning. *Journal of Turkish Science Education*, *17*(4), 520–527. https://doi.org/10.36681/tused.2020.42

Tyan, P. H., Rahman, F. A., & Sarvestani, M. S. (2020). Teachers' readiness in implementing and facilitating 21st century learning. *Universal Journal of Educational Research*, *8*(1 A), 24–29. https://doi.org/10.13189/ujer.2020.081304

Wahyu, Y., Suastra, I. W., Sadia, I. W., & Suarni, N. K. (2020). The Effectiveness of Mobile Augmented Reality Assisted STEM-Based Learning on Scientific Literacy and Students ' Achievement. *International Journal of Instruction*, *13*(3), 343–356.

Wulandari, S., Purdhiyah, P., & Annovasho, J. (2022). Studi Literatur : Penerapan E-Book Dalam Pembelajaran Fisika. *Jurnal Pendidikan Fisika Dan Sains*, *5*(02), 7–17.

Zuhdi, M., Gunada, I. W., & Taufik, M. (2024). Uji Instrumen Penguasaan Konsep Fisika Pada Pokok Bahasan Elastisitas Dan Hukum Hooke. *Kappa Journal*, *8*(1), 57–61.