Journal of Science and Mathematics Letters
ISSN 2462-2052 • eISSN 2600-8718

*Research Article*

# Psychometric Evaluation of a Digital Literacy Instrument Using the Rasch Rating Scale Model in the Context of Prospective Science Teachers

**Erna Noor Savitri[1], Sudarmin[2*], Murbangun Nuswowati[2], Sigit Saptono[3], Ani Rusilowati[1], Endang Susilaningsih[2], Roswanna Safkolam[4]**

[1] Science Education Doctoral Study Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, 50229 Semarang, Indonesia
[2] Chemistry Education Study Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, 50229 Semarang, Indonesia
[3] Biology Education Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, 50229 Semarang, Indonesia
[4] General Science Program, Faculty of Science Technology and Agriculture, Yala Rajabhat University, 95000 Yala Province, Thailand
* Corresponding author: sudarmin@mail.unnes.ac.id

**ARTICLE HISTORY**

**Received**
4 August 2025
**Revised**
2 October 2025
**Accepted**
25 November 2025
**Published**
2 January 2026

**KEYWORDS**

digital literacy
Rasch model
prospective science teacher

**ABSTRACT**

This research aims to develop a digital literacy instrument and test it on prospective science teachers using the Rasch Rating Scale Model. The study involved 81 second-semester students from the Science Education Study Program at Semarang State University (UNNES) and 59 students from UIN Salatiga. The instrument consists of 30 items, 10 essay questions, and 20 questionnaire statements measured using a four-point scale. Data were analyzed using the WINSTEPS Version 5.7.3.0 application. Content validity results show that all essay questions are valid, while 4 out of 20 questionnaire items are inconsistent and require revision. Construct validity analysis indicates that the essay questions account for 29% of the total variance, and the questionnaire explains 46%. The internal consistency (KR-20) is acceptable for essay questions (0.65) and very high for the questionnaire (0.83). Item reliability is excellent, with 0.92 for essay questions and 0.98 for the questionnaire. Item characteristics show a standard deviation of 0.35 for essay questions and 1.01 for the questionnaire. Reliability was evaluated using KR-20 for essay-based items and Cronbach's Alpha for the questionnaire. This ensures consistency in reporting reliability across both test and non-test components. Differential Item Functioning (DIF) analysis reveals bias in several essay questions between groups, with low probability values and high DIF contrast. In contrast, the questionnaire items function similarly across both groups. Biased items were reviewed and revised. Overall, the instrument demonstrates strong reliability and is suitable for measuring digital literacy in prospective science teachers.

https://doi.org/10.37134/jsml.vol14.1.8.2026

## 1. INTRODUCTION

Digital literacy is an individual's interest, attitudes, and abilities in using digital technology and communication tools to access, manage, integrate, analyze, and evaluate information, build new knowledge, and create and communicate with others in order to participate effectively in society (Li et al., 2021; Naamati-Schneider, 2024). Digital literacy is one of the essential competencies for facing the challenges of 21st-century learning. In the current digital era, science learning cannot be separated from the use of online resources, simulations, data analysis software, and collaborative platforms. Teachers are required not only to master scientific content but also to integrate technology effectively to foster students' higher-order thinking skills. Therefore, evaluating the digital literacy of prospective science teachers is particularly significant, since they will be at the frontline of preparing future generations who are scientifically literate and technologically adaptive.

According to Peng & Yu (2022), digital literacy is the ability to discover, evaluate, utilize, share and create ideas responsibly and ethically using tools and technology. Digital literacy is the ability possessed by a person to understand and utilize information originating from a very wide variety of sources in various formats accessed via a digital device (Chassignol et al., 2018; Giovanni & Komariah, 2019; OECD, 2019; Suherman, 2020). Digital literacy can also be interpreted as a person's ability to sort various information, interpret messages, and communicate effectively with other people (Restianty, 2018). According to Suherdi et al. (2021). Digital literacy is the ability to understand and use information widely and freely obtained through digital assistance. This broad and free use is, of course, still within the scope of norms, ethics, and culture. Digital literacy is also an indicator in education and culture to create critical and creative ways of thinking for students. Digital literacy is also considered to be the ability to train various abilities, competencies, and ICT (Information and Communication Technology) skills that are really needed by someone to adapt to the environment in the current technological era (Naila et al., 2021; Rusilowati et al., 2016; Techataweewan & Prasertsin, 2018). Digital literacy applied in various scopes, both in schools, families, and communities, is also considered capable of encouraging the progress of society in the 21st century era, like today (Abulibdeh et al., 2024; OECD, 2019; Sujana & Rachmatin, 2019).

While various terms such as information literacy, media literacy, and ICT literacy have been used in previous studies, this research adopts a more specific perspective. In this study, digital literacy is defined as the set of knowledge, skills, and attitudes that enable prospective science teachers to access, evaluate, create, and communicate information effectively through digital technologies within the context of science learning. This operational definition positions digital literacy as broader than ICT literacy, but more specific than general media or information literacy. The results of research on the theme of digital literacy have been carried out by several researchers, including research on increasing digital literacy that can be carried out using the Real Science Mask with QR Code media (Savitri et al., 2021). Apart from that, research also uses the Community Science Technology learning model to determine its effect on science learning outcomes in terms of digital literacy. The use of website-based learning media can also improve students' digital literacy skills in physics learning (Rahayu & Mayasari, 2018).

The development of the instrument was based on several theoretical considerations. UNESCO (2018) emphasizes digital literacy as a multidimensional construct involving technical, cognitive, and socio-emotional domains. According to Ng (2012), proposed cognitive, technical, and socio-cultural dimensions, while Eshet (2012) highlighted digital skills ranging from photo-visual to socio-emotional. Considering these perspectives, the current study integrates two frameworks that are particularly relevant to teacher education. Hague & Payton, (2022) proposed dimensions of digital literacy focusing on critical use of digital media, creativity, collaboration, and communication. Techataweewan & Prasertsin (2018) developed four main factors with twelve indicators of digital literacy that emphasize knowledge, skills, ethics, and usage. The integration of these two models was chosen because they combine socio-cultural and pedagogical aspects from Hague & Payton, with practical and technical indicators form Techataweewan & Prasertsin, thus providing a more comprehensive framework for assessing digital literacy among prospective science teachers.

Many recent studies also emphasize the importance of developing valid and reliable digital literacy instruments tailored to the educational context. Nguyen and Habók (2024) emphasized that teachers need digital literacy assessment tools that measure technical skills and pedagogical competencies in integrating technology into learning. At the student level, Avinç and Doğan (2024) developed a Digital Literacy Scale for secondary school students with validity and reliability tests using the Rasch model. Meanwhile, Son and Ha (2025) highlighted the need for contextualized digital literacy instruments, particularly in science practice, so that aspects of data collection, analysis, and communication can be measured more specifically. The findings from these three studies suggest that digital literacy measurements should not be generalized, but rather contextualized to the role (teacher/student) and the learning context (general or discipline-specific).

The measurement of digital literacy is explained in Digital Literacy Across the Curriculum (Hague & Payton, 2022), which explains eight dimensions of digital literacy, namely: Functional Skill and Beyond, Creativity, Collaboration, Communication, The Ability to find and select Information, Critical Thinking and Evaluation, Cultural and Social Understanding, and *E-Safety.* Literacy is not just reading or writing, but digital literacy is the knowledge and skills to use digital media, communication tools, or networks in finding, evaluating, using, creating information, and utilizing it in a healthy, wise, intelligent, careful, precise, and appropriate manner (Gonzalez-de-Eusebio & Tucho, 2021; Maulida & Sunarti, 2022). In Indonesia, there is not much research that examines digital literacy in science learning. Previous research has studied information literacy, media literacy, and ICT literacy in certain community groups in several regions of Indonesia. Digital literacy competencies are useful for dealing with information from various digital sources, which continues to develop along with the development of information and communication technology as a result of the media convergence phenomenon (Arfadila et al., 2022; Fitriarti, 2019; Yuniarto & Yudha, 2021).

The research problem raised in this study is the unavailability of valid, reliable, and structured indicator-based digital literacy assessment instruments in Indonesia's science education context. The need for such measuring instruments is critical to measure the digital competence of prospective science teachers accurately and systematically. The novelty of the research lies in the development of a digital literacy instrument based on the integration of indicators from two primary sources, namely Hague & Payton (2022) and Techataweewan & Prasertsin (2018) which have never been specifically adopted in the context of Indonesian science education, as well as the analysis of the validity and reliability of the instrument using the Rasch Rating Scale Model approach which can detect item quality and potential bias between respondent groups. The significance of the research will provide a real contribution to the development of digital literacy measuring instruments that are contextual, adaptive to technological developments, and appropriate for evaluating the readiness of prospective teachers in facing the challenges of 21st-century learning. Therefore, the study aims to develop and test the validity and reliability of a digital literacy instrument based on integrated indicators through the Rasch Rating Scale Model analysis on prospective science teachers in Indonesia.

According to the results of research on digital literacy indicators conducted by Techataweewan and Prasertsin (2018), the criteria for digital literacy in Thai undergraduate students consist of four factors containing 12 indicators, namely (1) Operational skills (cognition, discovery, and presentation); (2) thinking skills (analysis, evaluation, and creativity); (3) collaboration skills (teamwork, networking, and sharing); (4) awareness skills (ethics, legal literacy, and self-protection). Based on literacy indicators from Techataweewan & Prasertsin (2018) and Hague & Payton (2022), modifications are made by assimilating these two sources. In this study, these indicators were adapted and combined with Hague and Payton's framework to construct the initial blueprint of the instrument. Digital literacy indicators were used to develop a digital literacy measurement instrument in the form of 30 questions consisting of 10 essay questions and 20 questionnaires (non-test). Next, the instrument was analyzed for validity, reliability, and item analysis using the Rasch Rating Scale Model.

The Rasch Rating Scale Model (RSM) is a model in item response theory (IRT) that is used to analyze ordinal data, such as data derived from a Likert scale or similar rating scale (Azizah & Wahyuningsih, 2020; Jong et al., 2015; Tejada et al., 2011). RSM is a model in the Rasch family that allows for handling items with ordered (ordinal) response categories (Gómez et al., 2012; Quintão et al., 2013). Characteristics of the Rasch Rating Scale Model is designed to handle data with tiered rating scales, such as the Likert scale, which is often used in psychological and educational surveys (Morgan et al., 2022; Nguyen & Habók, 2024; Reddy et al., 2023). RSM assumes that all items in a scale have the same set of thresholds between categories, but item difficulty can be different. The advantages of the Rasch Rating Scale Model are: (1) Scale Consistency: provides a way to ensure that the ordinal scale functions consistently across all items; (2) Easy Interpretation: The item and person parameters in the Rasch model have an intuitive interpretation and are easy to understand; (3) DIF Detection: this model can be used to detect Differential Item Functioning (DIF) by adding group parameters and checking whether items function differently in different groups (Lestari et al., 2022; Maulana et al., 2023).

The background of the study is the need for a valid and reliable digital literacy assessment instrument in the context of science education. Previous research on digital literacy has primarily emphasized information and media components, lacking the development of structured, validated measurement tools. The increasing use of digital technology in education demands instruments capable of validly and reliably measuring digital literacy, particularly in science education. Technology-based learning, such as flipped classrooms, demands students' cognitive and technical readiness, closely related to digital literacy (Alias et al., 2021). This is increasingly relevant in the post-pandemic era, when online learning has become an integral part of the teaching and learning process, including in boarding schools, which face unique challenges regarding affordability and access to technology (Halim & Jamari, 2023)

Furthermore, implementing various digital tools in science learning in secondary schools, as found in the study by Mustapha & Jamaludin (2021), confirms that science teachers and prospective teachers need to be equipped with robust digital literacy skills to optimize technology in learning. However, few studies have explicitly developed and tested digital literacy measurement instruments, such as the Rasch Model, based on robust theory and psychometric analysis. Therefore, this study aims to develop a digital literacy instrument that suits the needs of 21st-century learning, and to test its validity and reliability using the Rasch Rating Scale Model approach. Therefore, developing this instrument is crucial to fill this gap and contribute to developing assessments that are in accordance with the needs of 21st-century learning, especially in the context of prospective science teachers. This study specifically aims to construct a digital literacy instrument based on integrated indicators from previous frameworks, test its content and construct validity using the Rasch Rating Scale Model, evaluate its reliability, analyze item difficulty levels, and identify any potential bias between respondent groups through DIF analysis.

## 2. METHODOLOGY

The research procedure consisted of (1) constructing the blueprint, (2) conducting expert review for content validity, (3) pilot testing the instrument with participants, and (4) analyzing the data using Rasch Rating Scale Model (RSM). In the initial stage, a literature review was carried out regarding digital literacy indicators. The instrument was developed based on a modification of the literacy indicator from Techataweewan & Prasertsin, (2018) and Hague & Payton (2022). Modification of the indicators was carried out by integrating indicators from Techataweewan & Prasertsin (2018), which emphasize operational skills, critical thinking, collaboration, and ethical awareness, as well as dimensions from Hague & Payton (2022), which include functional skills, collaboration, communication, and digital security. The integration process was done through content analysis to find intersections or substantial differences, which were then elaborated on in the item blueprint. After finding appropriate indicators, a blueprint item was created as a basis for creating 30 questions consisting of 10 essay questions (questions number 1 to 10) and the remaining 20 questionnaire questions (non-test).

This study employed a survey design with both test and non-test instruments, analyzed using the Rasch Rating Scale Model (Avinç & Doğan, 2024; Quintão et al., 2013; Son & Ha, 2025). The sampling technique used was purposive sampling, taking into account students who had taken basic digital literacy or educational technology courses. The number of items selected (30 items) took into account the coverage of eight dimensions of digital literacy indicators as well as the balance between test items (essays) and non-test items (questionnaires) to obtain a comprehensive picture. The participants were 140 prospective science teachers enrolled in teacher education programs at two universities in Indonesia. They were selected using purposive sampling, and the demographic characteristics such as age, gender, and study program were recorded. In this manuscript, the term participants is used consistently. These students are 81 second-semester students in the Science Education Study Program at Semarang State University and 59 students in the Science Education Study Program at UIN Salatiga.

The instrument was developed through several stages. First, a literature review was conducted to identify existing digital literacy frameworks. The initial pool of items was adapted from 54 items developed by Techataweewan & Prasertsin (2018) and several indicators from Hague & Payton (2022). After eliminating redundancies and aligning with the context of science education, 30 items were retained, consisting of 10 essay questions (test items) and 20 questionnaire items (non-test items) on the digital literacy measurement instrument, which is tested on students. Expert judgment was conducted with three experts in science education and educational technology to establish the instrument's content validity.

The questionnaire used four scales, namely SS (Strongly Agree), S (Agree), TS (Disagree), and STS (Strongly Disagree), with a score of 1 to 4, so that the data obtained was dichotomous data (Lafifa & Rosana, 2024). Data analysis was conducted using WINSTEPS Version 5.7.3.0 software. Rasch modeling provided item fit statistics, person reliability, item reliability, and principal component analysis (PCA) of residuals for construct validity. Differential Item Functioning (DIF) was also applied to examine potential bias across gender groups. This analysis will produce item parameters that fit the Rasch model in terms of validity, reliability, and item analysis data. In this case, there are three types of validity tests, namely empirical content validity tests and construct validity tests using Rasch. The finalized instrument consisted of 30 items, including 10 essay questions and 20 questionnaire items, which were mapped based on the developed digital literacy indicators. The distribution of these items according to each indicator and sub-indicator is presented in Table 1.

**Table 1.** Distribution of Question Items Based on the Development of Digital Literacy Indicators

| Indicator | Sub Indicator | Question Items |
|---|---|---|
| **Essay Questions (Test)** | | |
| Think creatively about creating products and models by utilizing digital technology | 1. Generating more than one idea and many ways to solve problems by using technology (fluency)<br>2. Look for many alternatives and different approaches to using technology (flexibility)<br>3. Able to produce new and unique products/thoughts (originality)<br>4. Able to detail the details of an object, idea, or situation so that it becomes more interesting (elaboration).<br>5. Able to define or understand in a different way from the usual way of using technology (redefinition) | 1,2,3,4,5 |
| Think critically in contributing and analyzing when dealing with information | 1. Able to formulate and analyze classification questions<br>2. Consider and observe sources of information<br>3. Make conclusions by considering the results of induction and deduction<br>4. Clarify terms in the information<br>5. Defend decisions by considering and thinking logically about premises, reasons, assumptions, positions, and other proposals | 6,7,8,9,10 |
| **Questionnaire (non test)** | | |
| Delivering product results by utilizing digital technology | 1. Able to find information through digital technology (information retrieval)<br>2. Able to read sources of scientific information (scientific reading) | 11, 12 |
| Communicating ideas or concepts through digital media | 1. Able to listen and observe through digital media (listening and observing)<br>2. Writing scientific ideas or ideas (scientific writing)<br>3. Able to present knowledge and information (knowledge presentation) | 13, 14, 15 |
| Discuss ideas and participate in digital spaces | 1. participate actively in the digital space<br>2. work productively discussing ideas or concepts with friends<br>3. responsible and flexible in group assignments<br>4. mutual respect between group members | 16, 17, 18, 19 |
| Assess and re-evaluate information found via the internet quickly and precisely | 1. Verify the facts and data presented.<br>2. Check the accuracy of the information through other independent sources.<br>3. Assess whether the information is supported by strong evidence.<br>4. Assess whether the information is relevant to the discussed topic or problem.<br>5. Ensure that the information is appropriate to the needs and context of its use. | 20, 21, 22, 23, 24 |
| Use of technology according to the context of social and cultural understanding | 1. Able to use technology according to local cultural norms, values , and habits<br>2. Able to sort information according to the context of social and cultural understanding<br>3. Preserving social culture through technology | 25, 26, 27 |
| Able to guarantee security when using digital technology | 1. Able to maintain data privacy and confidentiality in accordance with local ethical standards.<br>2. Be aware of personal information shared online and limit access to it.<br>3. Implement privacy on social media and other applications to minimize risks. | 28, 29, 30 |

The research instrument, which has developed indicators and questions, will then undergo content validity testing. The content validity test will involve five experts. The results of the expert validation will be analyzed using Aiken V. If the Aiken V analysis results indicate validity for the instrument being assessed, the instrument will then be tested. The results of the instrument test will be used to analyze construct validity, reliability, Item Difficulty Level, DIF (Intergroup Bias), Characteristics of Question Items, and Characteristics of Questionnaires using RASCH analysis. Construct validity is measured by referring to item polarity. Item polarity refers to the Point Measure Correlation (PTMEA Corr) value observed in the range of 0.20 to 0.71. In addition, construct validity is also seen through Principal Component Analysis (PCA), which examines items that correspond to the unidimensionality construct (Jin et al., 2020; Le et al., 2022). Reliability is analyzed based on person reliability and item reliability. The person reliability and item reliability categories use categories according to Boone (2016), namely <0.67 (weak); 0.67-0.80 (fair), 0.80-0.90 (good); 0.91-0.94 (excellent), >0.94 (Excellent). Cronbach's alpha is used to assess the internal consistency or reliability of a set of items in a test or questionnaire. Internal consistency refers to how well the items in an instrument measure the same concept. Person reliability and item reliability are two important aspects of measurement that are closely related to internal consistency as measured by Cronbach's Alpha. Cronbach's Alpha criteria can be seen in Table 2.

**Table 2.** Criteria for Cronbach's Alpha Value (Reliability) for Question Items

| Cronbach's Alpha Coefficient | Interpretation |
|---|---|
| < 0,50 | The scale has no internal consistency |
| 0,50 – 0,60 | The internal consistency of the scale is weak |
| 0,60 – 0,70 | The internal consistency of the scale is acceptable |
| 0,70 – 0,80 | The scale has internal consistency |
| >0,80 | The internal consistency of the scale is very high |

## 3. RESULTS AND DISCUSSION

Content validity was confirmed through expert review, which ensured that all items adequately represented the intended domain of digital literacy. Meanwhile, construct validity was examined using Rasch analysis, which provided evidence of item polarity, unidimensionality through PCA, and item fit indices. This distinction clarifies the methodological approach in evaluating the validity of the instrument. The results of the validity test showed that most of the questions were declared valid, while several items needed to be revised based on the Rasch Model fit criteria.

### *3.1. Content Validity*

Content validity was established through expert judgment to ensure that each item represented the intended construct of digital literacy. Content validity is conducted to ensure that the instrument items are in accordance with the indicators and constructs to be measured. In this study, content validity was assessed by five experts who are competent in their fields. Five experts participated in the validation process, consisting of two experts in science education, one expert in educational measurement and evaluation, and two experts in educational technology. The experts were asked to assess each instrument item using a specific scale. Each expert rated the relevance, clarity, and representativeness of the items using a four-point scale (1=not relevant to 4=highly relevant). The use of five experts/specialists aims to obtain a more objective assessment, reduce individual bias, and provide a stronger basis for determining the research instrument's feasibility. In the content validity analysis, the number of experts involved affects the results of Aiken's V calculation. The more experts there are, the more stable the Aiken's V value produced will be, and can describe a stronger level of agreement. The critical value (cut-off) used when using five experts with a scale of 1 to 4 is an Aiken's V value ≥ 0.80, then the item is considered valid. If the Aiken's V value < 0.80, the item must be revised or removed. The results of content validity by experts are presented in Table 3.

**Table 3.** Aiken V Analysis of Content Validity by Experts Judgement

| Validation Items | Validator | | | | | ∑s | n(C-1) | V | Criteria |
|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | | | | |
| 1 | 4 | 3 | 3 | 4 | 4 | 13 | 15 | 0.87 | Highly valid |
| 2 | 4 | 4 | 4 | 4 | 4 | 15 | 15 | 1.00 | Highly valid |
| 3 | 4 | 4 | 4 | 3 | 4 | 14 | 15 | 0.93 | Highly valid |
| 4 | 4 | 4 | 3 | 4 | 4 | 14 | 15 | 0.93 | Highly valid |
| 5 | 4 | 4 | 3 | 4 | 4 | 14 | 15 | 0.93 | Highly valid |
| **Average V value** | | | | | | | | **0.93** | **Highly valid** |

Table 3 presents a summary of Aiken's V results, which show that all five sample items achieved coefficients above 0.80 with an average of 0.93, confirming that the instrument's content aligns with the theoretical framework of digital literacy. Therefore, the expert assessment results support the content validity of the developed instrument. Further analysis using the Rasch Rating Scale Model was conducted separately to examine construct validity, reliability, and item characteristics. The results of the focus group discussion with experts supported the item revisions, particularly in terms of clarity of wording and contextual alignment with science teacher education. The Focus Group Discussions (FGDs) feedback was integrated into the final version of the instrument before Rasch analysis.

### *3.2. Construct Validity, Reliability (Cronbach's Alpha), Item Difficulty Level*

The instrument, which was declared content-valid through Aiken's V analysis, was then pilot-tested to obtain empirical data. The pilot-test data were used to analyze construct validity, reliability, and item difficulty. Construct validity testing was conducted to ensure the instrument items accurately measured the intended theoretical constructs. Instrument reliability was analyzed using Cronbach's Alpha coefficient to assess internal consistency between items. Meanwhile, difficulty analysis was used to determine the extent to which the instrument items were considered easy, moderate, or difficult for respondents. Construct Validity, Reliability (Cronbach's Alpha), and Item Difficulty Level are presented in Table 4.

**Table 4.** Results of Validity, Reliability, and Item Difficulty Level

| Analysis Aspect | Instrument Type | Main Findings | Revised Items | Valid Items |
|---|---|---|---|---|
| **Construct Validity** | Essay | All questions are still considered fit, although some need major revision<br>Explained variance is only 29% → low validity | Questions 7, 9 (2 requirements not met)<br>Revision of the model/instrument is recommended | 10 |
| | Questionnaire | 4 items are invalid (16, 18, 19, 20) because the MNSQ, ZSTD, and Pt. Measure Corr values are outside the limits<br>Explained variance 46% → quite good validity | 16, 18, 19, 20<br>Items with high contrast (e.g., item 4) need to be reviewed | 16 |
| **Reliability (Cronbach's Alpha)** | Essay | 0.65 → Quite good (acceptable) | - | - |
| | Questionnaire | 0.83 → Very high | - | - |
| **Item Difficulty Level** | Essay | SD = 0.35. Most questions are classified as moderate | Question 3 (difficult), question 6 (easy) | - |
| | Questionnaire | SD = 1.01. High variation: items 1, 4, 6 are very difficult; items 3, 8, 14 are very easy | 1, 4, 6 | - |

The results of the construct validity analysis using Principal Component Analysis (PCA) showed that the essay questions could only explain 29% of the total variance. This relatively low percentage indicates that the construct validity is still weak, and the questions do not fully represent the dimensions of digital literacy to be measured. This can be caused by several factors, such as the inconsistency between the questions and indicators, the disproportionate coverage of the indicators, and the formulation of questions that are too complex or have multiple interpretations for the cognitive level of the respondents. In addition, unexpected latent dimensions or overlaps between constructs can also cause distortion of the primary measurement focus. To overcome these problems, it is necessary to review the relationship between each question item and the referred indicator, improve the wording of the questions to be more precise and more focused, and develop a more comprehensive digital literacy model to suit the competencies to be measured. Follow-up actions that can be taken include revising or replacing weak items, involving material experts and evaluation in the validation process, and conducting cognitive interviews or response process analysis to identify and reduce variants that are not relevant to the construct. Retesting on a larger sample is also recommended to verify the improvement in construct validity after the revision. Meanwhile, the questionnaire components showed better construct validity, with 46% of the variance explained by the primary construct. This indicates that the items in the questionnaire are more aligned with the intended digital literacy construct. However, there were still unexplained variances and high contrasts in the residual components, indicating the possibility of unexpected latent dimensions. This could be because some statement items were too general, not contextually specific, or had a certain cultural content that all respondents did not uniformly understand. Suggested solutions include identifying and improving items that contribute to high contrast, simplifying the language, and adjusting the statements' context to suit the respondents' backgrounds and experiences. Important follow-up actions that need to be carried out include conducting exploratory factor analysis (EFA) to verify the dimensional structure of the questionnaire, focus group discussions (FGDs) to obtain qualitative feedback, and refining the conceptual framework to make it more straightforward and more operational. With these steps, this instrument can have stronger construct validity and accurately measure prospective science teachers' digital literacy competencies accurately and fairly.

The reliability analysis results show that both the essay and questionnaire scales have a level of reliability that is classified as sufficient to very high. The Cronbach's Alpha value for the essay questions is 0.65, indicating acceptable internal consistency, while for the questionnaire, it reaches a value of 0.83, which is categorized as very high. In addition, the item reliability value reaches 0.92 for the essay and 0.98 for the questionnaire, indicating that both parts of the instrument have excellent measurement stability. Further reliability analysis at the construct level (knowledge, skills, ethics, and usage) showed coefficients ranging from 0.79 to 0.85, indicating acceptable internal consistency within each domain. These results indicate that, in general, the instrument developed has strong internal consistency, especially in the questionnaire section. The high reliability of the questionnaire can be caused by the closed and more structured scale format, so that respondents provide stable and consistent answers.

On the other hand, the slightly lower reliability of the essay is likely influenced by variations in respondents' understanding of the questions, writing skills, and differences in interpretation of the indicators being measured. In addition, open-ended questions, such as essays, are more susceptible to subjectivity in assessment, especially if they are not equipped with a detailed assessment rubric. Solutions that can be applied to improve reliability in the essay section are to improve the assessment rubric to be more objective and standardized, and to train assessors to ensure uniformity in the scoring process. In addition, the formulation of essay questions also needs to be simplified and focused to be more specific to the intended indicators to reduce the variability of irrelevant responses. Follow-up actions that need to be carried out are inter-rater reliability testing to ensure consistency between assessors and re-testing the reliability of the revised version of the questions using a more diverse sample. Although the reliability of the questionnaire section is very high, it is still necessary to pay attention to the potential for redundancy between items. Therefore, conducting an inter-item correlation analysis is recommended to identify items that are too similar and simplify items if necessary so that the instrument remains efficient without sacrificing its accuracy.

Item-difficulty analysis indicated a balanced distribution of logit values. Essay items had a standard deviation of 0.35, categorizing most questions as moderate in difficulty, while questionnaire items had a wider spread (SD = 1.01), indicating greater variation in perceived difficulty. Items that were extremely easy or difficult were identified for potential revision to maintain optimal discrimination across respondent abilities. Such items can compromise the instrument's effectiveness in distinguishing respondents' actual ability levels. Items that are too easy tend to fail to identify respondents with low ability, while items that are too difficult risk eliciting random responses from respondents. Therefore, we reviewed the extreme items based on logit values and content validity, then restructured the items to align with students' cognitive developmental levels and contextualize their experiences. Items that were too difficult were simplified and divided into two levels, while too easy items were expanded to become more challenging. The distribution of item difficulty levels mirrors the indicator progression suggested by Techataweewan & Prasertsin (2018), ranging from basic information access to high-level problem-solving and innovation. This theoretical alignment strengthens the argument that the developed instrument is conceptually grounded and psychometrically robust.

### *3.3. DIF (Intergroup Bias)*

DIF analysis was conducted using the Rasch Rating Scale Model. DIF detects whether individuals from different groups but with the same ability level have different probabilities of endorsing or correctly responding to an item (Boone, 2016). In this study, DIF was analyzed by comparing responses between students from Universitas Negeri Semarang (UNNES) and UIN Salatiga. DIF Analysis presented in Table 5.

**Table 5.** DIF Analysis

| Analysis Aspect | Instrument Type | Main Findings | Revised Items | Valid Items |
|---|---|---|---|---|
| **DIF (Intergroup Bias)** | Essay | Items 1, 3, 4, 5, 6, 7, 9 show significant differences between UIN Salatiga & UNNES | 7 items need to be evaluated | - |
| | Questionnaire | Most items do not show significant DIF, except items 4 and 16 | 4, 16 | - |

The solution to overcome DIF findings is to revise the items that are identified as biased, either by changing the wording to be more neutral and universal or by adding an equal context for all groups. It is also recommended that a cultural fairness test be conducted and reviewers from across institutions be involved in compiling the questions. Follow-up actions that need to be taken include retesting the instrument after revision with a broader and more diverse group of respondents, as well as the application of advanced statistical tests such as Mantel-Haenszel and Rasch-Welch t to ensure that all items function fairly for all respondents. Several digital literacy question instrument items showed significant bias between UIN Salatiga and UNNES. This is indicated by very low probability values and considerable DIF contrasts. Items with significant differences were then reviewed and revised to ensure they were not biased toward one group. Identification of biased items aims to improve the validity and fairness of the assessment instrument. The following analysis was carried out to find the DIF (Differential Item Functioning) t value. This analysis detects the presence of DIF in test items or questionnaires. T value in this context refers to the value obtained from a statistical test comparing item performance between two groups.

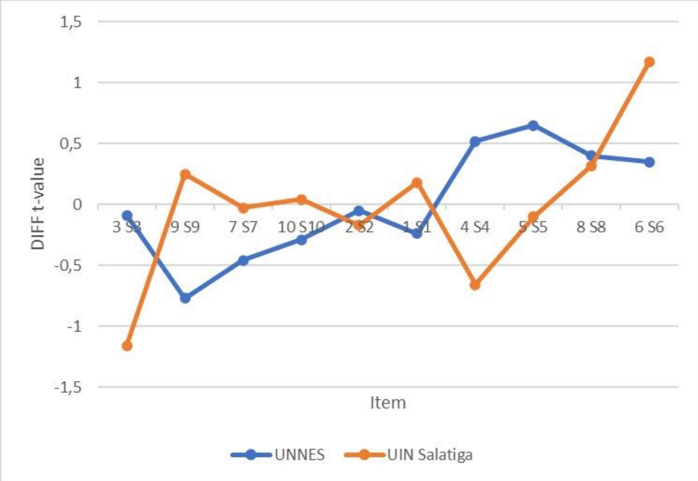


**Figure 1.** DIF t-Value on Digital Literacy Questions

Based on Figure 1, it can be seen that items with very high or very low t-values indicate that there is a significant difference between UIN Salatiga and UNNES for the item. These items require further evaluation to determine whether the difference is due to bias in the item or legitimate differences in group abilities. Items that show significant differences require revision to ensure that the items are fair and not biased towards one group. Items with t-values close to zero indicate that there is no significant difference between UIN Salatiga and UNNES for the item. These items tend to be fair and not biased between groups. The results of the DIF analysis, Direction of Bias and Recommendations are presented in Table 6.

**Table 6.** Results of DIF analysis, Bias Direction and Recommendations

| Instrument | Number of Items Bias (Significant, $p < 0.05$) | Item DIF | Direction of Bias | Recommendation |
|---|---|---|---|---|
| Essay | 7 out of 10 questions | Questions 1, 3, 4, 5, 6, 7, 9 | Mixed (some are harder for UNNES, others for UIN) | Revise or recalibrate questions that show bias |
| Questionnaire | 2 out of 20 items | Items 4 and 16 | Items are harder for certain groups (not specifically mentioned) | Evaluate context and item formulation |

### *3.4. Characteristics of Question Items*

The characteristics of the analyzed items are the level of difficulty of the items using Item Measure on the WINSTEPS Version 5.7.3.0 application. The standard deviation (SD) value in this test is 0.35, so the category of the Level of Difficulty of the Items is based on the Logit Value in Table 8. Items with unsatisfactory fit should be reviewed because they do not measure what they are supposed to measure or may confuse respondents. This is very important to ensure that the measurement of the desired ability in psychometrics or education is valid and reliable. The results of the analysis of the level of difficulty of the test items based on the logit value and its category are presented in Table 7.

**Table 7.** Categories of Difficulty Levels in Digital Literacy Question Instruments

| Logit Value | Categories |
|---|---|
| Greater than +0.35 SD | Very Hard |
| 0.0 logit + 0.35 SD | Hard |
| 0.0 logit – 0.35 SD | Medium |
| Less than -0.35 SD | Easy |

Items with unsatisfactory fit should be reviewed because they do not measure what they are supposed to measure or may confuse respondents. This is very important to ensure that the measurement of the desired ability in psychometrics or education is valid and reliable. The results of the analysis of the level of difficulty of the test items based on the logit value and its category are presented in Table 8. Distribution of the difficulty level of digital literacy questions based on the logit value (measure) from the Rasch analysis. The logit value indicates each question's difficulty level, where a positive value indicates an easier question and a negative value indicates a more difficult question. The difficulty level categories are classified into three: easy (logit > +0.35), moderate (logit between -0.35 and +0.35), and difficult (logit < -0.35), with a standard deviation of 0.35 as the limit reference. Based on the results of the data analysis, most of the questions are in the moderate category, which is 8 out of 10. This shows that, in general, essay questions are in a proportional difficulty range and can measure respondents' abilities well. One question (question number 6) is in the easy category with a logit of +0.70, which indicates that most respondents can answer this question correctly. Meanwhile, only one question is classified as difficult, namely question number 3, with a logit value of -0.49. The items' distribution based on the difficulty level of the questions is presented in Table 9.

**Table 8.** Analysis of the Level of Difficulty of Question Items Based on Logit Values and Categories

| Instrument | Extreme Items (Most Difficult) | The Easiest Item | Distribution of Difficulty Levels | Recomendation |
|---|---|---|---|---|
| Essay (Question) | Question no. 3 (-0.49) | Question no. 6 (+0.70) | The majority are in the moderate category (7 questions), 1 difficult, 1 easy | Logit distribution is normal, varied, and needs a balance of difficulty level |
| Questionnaire | Item 4 (+2.34), 6 (+1.85), 1 (+1.40) → *Very Difficult* | Item 14 (-1.95), 8 (-1.26), 3 (-1.12) → *Easy* | 3 very difficult, 6 difficult, 9 moderate, 3 easy | Logit variation is high (SD = 1.01), needs revision of extreme items |

**Table 9.** Distribution of Items Based on the Level of Difficulty of Digital Literacy Questions

| Question Number | Total Score | Total Count | Measure | Question Type |
|---|---|---|---|---|
| 6 | 1061 | 140 | 0.70 | Easy |
| 8 | 1095 | 140 | 0.37 | Medium |
| 5 | 1099 | 140 | 0.33 | Medium |
| 4 | 1129 | 140 | 0.03 | Medium |
| 1 | 1138 | 140 | -0.06 | Medium |
| 2 | 1142 | 140 | -0.10 | Medium |
| 10 | 1147 | 140 | -0.15 | Medium |
| 7 | 1160 | 140 | -0.28 | Medium |
| 9 | 1167 | 140 | -0.35 | Difficult |
| 3 | 1182 | 140 | -0.49 | Easy |

The item distribution in Table 8 indicates that the instrument has a relatively balanced distribution of difficulty levels, although more attention needs to be paid to questions with extreme levels of difficulty. Question number 6, which is too easy, is at low risk of discriminating between respondents with low and high abilities, so it needs to be reviewed or further developed to be challenging. Conversely, question number 3, which is considered difficult, needs to be analyzed in terms of wording, complexity, or indicator suitability so as not to disproportionately hinder respondents' understanding. Suggested follow-up actions include revising these extreme questions and conducting additional trials to validate the effectiveness of the items in optimally measuring digital literacy skills.

### *3.5. Characteristics of Questionnaires*

The Standard Deviation (SD) value in the questionnaire instrument test is 1.01, so the category of Questionnaire Item Difficulty Level is based on the Logit Value in Table 10. The level of difficulty in this questionnaire can be said that the statements in the questionnaire can confuse respondents. Distribution of the difficulty level of digital literacy questionnaire items based on the logit value from the Rasch analysis results. This logit value describes the extent to which the item is considered easy or difficult by respondents, namely, prospective science teacher students. Based on the size and difficulty of prospective science teacher students, the distribution of questionnaire items is presented in Table 11.

Based on Table 11, it can be seen that most of the questionnaire items (13 out of 20 items) are in the moderate category, indicating that the majority of statements in the questionnaire have a balanced level of difficulty and are able to detect variations in respondents' abilities in general. Three items are classified as easy, namely items 3, 8, and 14, with logits below -1.00 each. These three items tend to be answered easily by all respondents, so their discrimination values need to be reviewed to remain informative. Meanwhile, four items are classified as difficult, namely items 7, 10, 11, and 18, which are slightly above logit 0 and approaching the upper limit. In addition, there are three items that are very difficult, namely item 1 (logit 1.40), item 6 (logit 1.85), and item 4 (logit 2.34). These items indicate that only respondents with very high abilities tend to answer them consistently. Very high levels of difficulty can cause unbalanced cognitive pressure and risk disrupting the validity of the measurement if not balanced by representative items. The variation in levels of difficulty found reflects that the questionnaire instrument has a wide range, but items with extreme logits (either too low or too high) need to be reviewed. This is important to ensure that all items contribute to measuring the digital literacy construct in a balanced manner. Follow-up actions that can be taken include re-analyzing the interpretation of items by respondents, revising the wording of extreme items, and conducting additional trials to reassess the diagnostic function of each item in a wider and more diverse population. The distribution of items based on the level of difficulty of the digital literacy questionnaire items can be presented in Figure 2. The moderate category dominates the distribution with a total of 8 items, indicating that most of the statements in the questionnaire have a proportional level of difficulty and are within the average ability of the respondents. The difficult category is in second place with six items, indicating that a number of statements are quite challenging for most prospective science teacher students. Meanwhile, the easy and difficult categories each consist of 3 items. The presence of items with extreme levels of difficulty like this (either too easy or challenging) needs to be explicitly considered because it risks reducing the instrument's effectiveness in optimally differentiating the respondents' abilities.

**Table 10.** Categories of Difficulty Levels in the Digital Literacy Questionnaire Instrument

| Logit Value | Categories |
|---|---|
| Greater than +1.01 SD | Very Hard |
| 0.0 logit + 1.01 SD | Hard |
| 0.0 logit – 1.01 SD | Medium |
| Less than -1.01 SD | Easy |

**Table 11.** Distribution of Items Based on the Level of Difficulty of Digital Literacy Questionnaire Items

| Question Number | Measure | Question Type |
|---|---|---|
| 3 | -1.12 | Easy |
| 8 | -1.26 | Easy |
| 14 | -1.95 | Easy |
| 20 | -0.12 | Medium |
| 13 | -0.15 | Medium |
| 2 | -0.18 | Medium |
| 19 | -0.21 | Medium |
| 12 | -0.27 | Medium |
| 5 | -0.49 | Medium |
| 9 | -0.81 | Medium |
| 15 | -0.81 | Medium |
| **7** | 0.93 | Difficult |
| 17 | 0.34 | Difficult |
| 16 | 0.32 | Difficult |
| 18 | 0.12 | Difficult |
| 11 | 0.05 | Difficult |
| 10 | 0.02 | Very Difficult |
| 4 | 2.34 | Very Difficult |
| 6 | 1.85 | Very Difficult |
| 1 | 1.40 | Very Difficult |

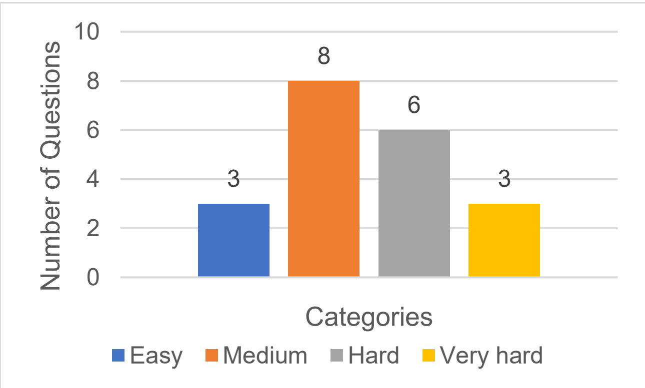


**Figure 2.** Distribution of Items Based on the Level of Difficulty of Digital Literacy Questionnaire Items

Overall, this distribution shows that the questionnaire instrument has covered a good variation in difficulty levels. However, further studies are still needed on items in the extreme category to ensure that the measurement quality remains valid, reliable, and balanced. This study has several limitations that need to be considered. First, the number of samples is limited to two institutions that may not represent the diversity of educational contexts in Indonesia. Second, although the instrument's reliability is high, the construct validity, especially in essay questions, is still low, so improving the model or expanding the indicators is necessary. Third, the potential bias of student responses due to an understanding of local socio-cultural contexts has also not been fully explored. Recommendations for further research include expanding the sample from multiple institutions, utilizing additional exploratory factor analysis to increase construct validity, and adjusting the cultural context in items to reduce potential bias. The results of this study are consistent with recent empirical evidence. Son and Ha (2025) reported that Rasch modeling provides useful diagnostic information about the reliability and dimensionality of digital literacy measures, similar to the results obtained here. In addition, Avinç and Doğan (2024) highlighted the importance of multidimensional perspectives in assessing pre-service teachers' digital competencies, which supports the integrated framework used in this study.

## 4. CONCLUSION

This study developed and validated a digital literacy instrument consisting of essay questions and questionnaires, tested on prospective science teacher students. The Rasch Rating Scale Model was used for analysis. The instrument showed good content validity for the essay items, though four questionnaire statements required revision. Construct validity results indicated moderate explanatory power (29% for essay items and 46% for questionnaires). Reliability analysis showed acceptable internal consistency (KR-20 = 0.65 for essays, 0.83 for questionnaires) and high item reliability (0.92 and 0.98, respectively). DIF analysis revealed item bias in some essay questions between student groups, while the questionnaire showed no significant bias. These findings highlight the instrument's strong potential, with further improvement needed in essay item construction and bias reduction. The revised instrument can serve as a reliable tool for measuring digital literacy in science teacher education. However, the relatively small sample size (N=140) and limited diversity of participants may have influenced the DIF results and generalizability of the findings. Future studies should involve larger and more diverse samples of prospective teachers.

**ACKNOWLEDGEMENT**
The authors would like to express their deepest gratitude to several important organizations for their significant support. We are especially grateful to the Higher Education Financing Agency (BPPT) for the scholarship provided through the Indonesian Education Scholarship Program (BPI). We also appreciate the contribution of Universitas Negeri Semarang, which has greatly enriched our research efforts.

**CONFLICT OF INTEREST**
The author declares no conflict of interest.

**AUTHOR CONTRIBUTION**
Erna Noor Savitri, Sudarmin: Conceptualization, Methodology, Data Curation, Software, and writing the original draft. Murbangun Nuswowati, Sigit Saptono: Data curation, Visualization, Investigation. Ani Rusilowati, Endang Susilaningsih: Supervision, Reviewing, and Editing.

**DATA AVAILABILITY**
The authors confirm that the data supporting the findings of this study are available within the article.

**DECLARATION OF GENERATIVE AI**
While preparing this work, the author used the Grammarly application to improve the readability and language of the manuscript. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

**ETHICS**
Not applicable.

## REFERENCES

Abulibdeh A, Zaidan E, Abulibdeh R. (2024). Navigating the confluence of artificial intelligence and education for sustainable development in the era of industry 4.0: Challenges, opportunities, and ethical dimensions. *Journal of Cleaner Production*, 437, 140527. doi:10.1016/j.jclepro.2023.140527

Alias Z, Mohamad Hussin MZ, Badruldin NA, Taha H. (2021). The effectiveness of flipped classroom method on achievement and readiness in sets topic for form 4 students. *Journal of Science and Mathematics Letters*, 9(2), 62-74. doi:10.37134/jsml.vol9.2.7.2021

Arfadila A, Aulia ERN, Nugraha RW, Humaeroh S. (2022). Penerapan E-learning dalam inovasi pendidikan untuk meningkatkan kemampuan literasi digital siswa. *Jurnal Pendidikan Sosiologi Dan Humaniora*, 13(2), 392. doi:10.26418/j-psh.v13i2.54771

Avinç E, Doğan F. (2024). Digital literacy scale: Validity and reliability study with the rasch model. *Education and Information Technologies*, 29(17), 22895-22941. doi:10.1007/s10639-024-12662-7

Azizah A, Wahyuningsih S. (2020). Penggunaan model Rasch untuk analisis instrumen tes pada mata kuliah Matematika Aktuaria. *JUPITEK: Jurnal Pendidikan Matematika*, *3*(1), 45-50. doi:10.30598/jupitekvol3iss1pp45-50

Boone WJ. (2016). Rasch analysis for instrument development: Why,when,and how? *CBE Life Sciences Education*, 15(4), 4. doi:10.1187/cbe.16-04-0148

Chassignol M, Khoroshavin A, Klimova A, Bilyatdinova A. (2018). Artificial intelligence trends in education: A narrative overview. *Procedia Computer Science*, 136, 16-24. doi:10.1016/j.procs.2018.08.233

Eshet Y. (2012). Thinking in the digital era: A revised model for digital literacy. *Issues in Informing Science and Information Technology*, 9(2), 267-276. doi:10.28945/1621

Fitriarti EA. (2019). Urgensi literasi digital dalam menangkal hoax informasi kesehatan di era digital. *Metacommunication; Journal of Communication Studies*, 4(2), 234-246. doi:10.20527/mc.v4i2.6929

Giovanni F, Komariah N. (2019). Hubungan antara literasi digital dengan prestasi belajar siswa SMA Negeri 6 Kota Bogor. *Libraria:*

*Jurnal Perpustakaan*, 7(1), 147-162. doi:10.21043/libraria.v7i1.5827

Gómez LE, Arias B, Verdugo MÁ, Navas P. (2012). Application of the Rasch rating scale model to the assessment of quality of life of persons with intellectual disability. *Journal of Intellectual and Developmental Disability*, 37(2), 141-150. doi:10.3109/13668250.2012.682647

Hague C, Payton S. (2010). Digital literacy across the curriculum. Bristol: Futurelab.

Halim AD, Jamari D. (2023). Exploring online learning experiences among students in full boarding school during covid-19 pandemic. *Journal of Science and Mathematics Letters*, 11, 103-109. doi:10.37134/jsml.vol11.sp.11.2023

Jin KY, Reichert F, Cagasan Jr LP, de La Torre J, Law N. (2020). Measuring digital literacy across three age cohorts: Exploring test dimensionality and performance differences. *Computers & Education*, 157, 103968. doi:10.1016/j.compedu.2020.103968

Jong C, Hodges TE, Royal KD, Welder RM. (2015). Instruments to measure elementary preservice teachers' conceptions: An application of the Rasch Rating Scale Model. *Educational Research Quarterly*, 39(1), 21-48.

Lafifa F, Rosana D. (2024). Innovations in assessing students'digital literacy skills in learning science: Effective multiple choice closed-ended tests using RASCH model. *Turkish Online Journal of Distance Education*, 25(3), 44-56.

Le B, Lawrie GA, Wang JTH. (2022). Student self-perception on digital literacy in STEM blended learning environments. *Journal of Science Education and Technology*, 31(3), 303-321. doi:10.1007/s10956-022-09956-1

Lestari W, Wigati I, Sholeh MI, Pramita D. (2022). Instrumen literasi digital guru menggunakan Model Rasch. *Orbital: Jurnal Pendidikan Kimia*, 6(2), 104-113. doi:10.19109/ojpk.v6i2.15019

Li J, Brar A, Roihan N. (2021). The use of digital technology to enhance language and literacy skills for Indigenous people: A systematic literature review. *Computers and Education Open*, 2, 100035. doi:10.1016/j.caeo.2021.100035

Maulana S, Rusilowati A, Nugroho SE, Susilaningsih E. (2023). Implementasi Rasch Model dalam pengembangan instrumen tes diagnostik. *Prosiding Seminar Nasional Pascasarjana*, 6(1), 748-757.

Maulida F, Sunarti T. (2022). Pengembangan instrumen tes literasi sains berbasis kearifan lokal di kabupaten lamongan. *ORBITA: Jurnal Pendidikan dan Ilmu Fisika*, 8(1), 52-65. doi:10.31764/orbita.v8i1.8337

Morgan A, Sibson R, Jackson D. (2022). Digital demand and digital deficit: conceptualising digital literacy and gauging proficiency among higher education students. *Journal of Higher Education Policy and Management*, 44(3), 258-275. doi:10.1080/1360080X.2022.2030275

Mustapha MB, Jamaludin AAB. (2021). Analysis of interest towards science subject and its relation with 21st century skills. *Journal of Science and Mathematics Letters*, 9(2), 103-116. doi:10.37134/jsml.vol9.2.10.2021

Naamati-Schneider L. (2024). Enhancing AI competence in health management: students' experiences with ChatGPT as a learning Tool. *BMC Medical Education*, 24(1), 1–16. doi:10.1186/s12909-024-05595-9

Naila I, Ridlwan M, Haq MA. (2021). Literasi digital bagi guru dan siswa sekolah dasar: Analisis konten dalam pembelajaran. *Jurnal Review Pendidikan Dasar: Jurnal Kajian Pendidikan Dan Hasil Penelitian*, *7*(2), 166-122. doi:10.26740/jrpd.v7n2.p166-122

Ng W. (2012). Can we teach digital natives digital literacy? *Computers & Education*, 59(3), 1065-1078. doi:10.1016/j.compedu.2012.04.016

Nguyen LAT, Habók A. (2024). Tools for assessing teacher digital literacy: a review. *Journal of Computers in Education*, 11(1), 305-346. doi:10.1007/s40692-022-00257-5

OECD. (2019). An OECD Learning Framework 2030. Springer.

Peng D, Yu Z. (2022). A literature review of digital literacy over two decades. *Education Research International*, 1, 2533413. doi:10.1155/2022/2533413

Quintão S, Delgado AR, Prieto G. (2013). Validity study of the beck anxiety inventory (Portuguese version) by the rasch rating scale model. *Psicologia: Reflexao e Critica*, *26*(2), 305–310. doi:10.1590/S0102-79722013000200010

Rahayu T, Mayasari T. (2018). Profil kemampuan awal literasi digital dalam pembelajaran fisika siswa SMK Kota Madiun. In *Seminar Nasional Quantum*, 25(1), 2477-1511.

Reddy P, Chaudhary K, Sharma B, Hussein S. (2023). Essaying the design, development and validation processes of a new digital literacy scale. *Online Information Review*, 47(2), 371-397. doi:10.1108/OIR-10-2021-0532

Restianty A. (2018). Literasi digital, sebuah tantangan baru dalam literasi media. *Gunahumas*, 1(1), 72-87. doi:10.17509/ghm.v1i1.28380

Rusilowati A, Kurniawati L, Nugroho SE, Widiyatmoko A. (2016). Developing an instrument of scientific literacy asessment on the cycle theme. *International Journal of Environmental and Science Education*, 11(12), 5718-5727.

Savitri EN, Dewi NR, Amalia AV, Rahmadani OEP, Kholidah A. (2021). Real science mask with QR code to improve the digital literation. *Unnes Science Education Journal*, 10(1), 34-40. doi:10.15294/usej.v10i1.42537

Son M, Ha M. (2025). Development of a digital literacy measurement tool for middle and high school students in the context of scientific practice. *Education and Information Technologies*, 30(4), 4583-4606. doi:10.1007/s10639-024-12999-z

Suherdi D, Rezky SF, Apdillah D, Sinuraya J, Sahputra A, Syahputra D, Wahyuni D. (2021). Peran literasi digital di masa pandemik. Deli Serdang-Indonesia: Cattleya Darmaya Fortuna Publisher.

Suherman A. (2020). Promoting digital literacy skills: An action research to people of Kampung Literasi. *Universal Journal of Educational Research*, 8(4), 1372-1386. doi:10.13189/ujer.2020.080430

Sujana A, Rachmatin D. (2019). Literasi digital abad 21 bagi mahasiswa PGSD: apa, mengapa, dan bagaimana. In *Current Research in Education: Conference Series Journal*, 1(1), 3-13.

Techataweewan W, Prasertsin U. (2018). Development of digital literacy indicators for Thai undergraduate students using mixed method research. *Kasetsart Journal of Social Sciences*, 39(2), 215-221. doi:10.1016/j.kjss.2017.07.001

Tejada AJR, Luque MN, Rojas OML, Moreno PJP. (2011). Prejudiced attitude measurement using the rasch rating scale model. *Psychological Reports*, 109(2), 553-572. doi:10.2466/07.17.PR0.109.5.553-572

Yuniarto B, Yudha RP. (2021). Literasi digital sebagai penguatan pendidikan karakter menuju era society 5.0. *Edueksos Jurnal Pendidikan Sosial & Ekonomi*, 10(2), 176-194. doi:10.37329/cetta.v6i3.2603