ELEMENTARY *Islamic Teacher Journal*
E-ISSN : 2503-0256 / ISSN : 2355-0155
Volume 13 Number 1 January - June 2025 (PP. 65-80)
http://dx.doi.org/10.21043/elementary.v13i1.33527
accessed at : http://journal.iainkudus.ac.id/index.php/elementary

Psychometric Analysis of Science Test Items and Their Implications for Scientific Literacy at Madrasah Ibtidaiyah

Ahmad Syaripudin[1*], Shaleh[2], Sihono[3]
[1-3]UIN Sunan Kalijaga Yogyakarta, Indonesia
**Corresponding author : 24204081005@student.uin-suka.ac.id*

Abstract

This study aims to analyze the quality of science test items developed by Madrasah Ibtidaiyah teachers and their implications for students' scientific literacy. The evaluation focuses on how well these items meet basic psychometric criteria—such as content validity, item difficulty, discrimination index, and distractor effectiveness—and their alignment with scientific literacy indicators used in the national AKMI assessment. A descriptive evaluative quantitative approach was employed, with data consisting of 30 multiple-choice questions and corresponding student literacy scores. The analysis reveals that while most items technically fulfill basic assessment standards, there is a noticeable gap between students' school-based test scores and their performance in national literacy assessments. This suggests that many test items fail to measure students' scientific reasoning and real-world application skills. Some items integrate higher-order thinking skills (HOTS), but the use of such items is inconsistent and lacks emphasis on contextual and reasoning-based dimensions. These findings indicate the need to shift from memorization-based evaluation models toward literacy-oriented, competency-based assessments. Strengthening teacher capacity in designing context-rich, curriculum-aligned test items is crucial. Effective assessments should go beyond measuring academic outcomes and actively foster students' scientific thinking, reasoning, and problem-solving skills, preparing them for real-world challenges in accordance with global assessment standards.

Keywords: *Akmi, Higher-Order Thinking Questions, Psychometrics, Scientific Literacy.*

INTRODUCTION

Science education in elementary schools is an important foundation in building science literacy from an early age (Febriandani et al., 2025). Science literacy includes not only mastery of concepts, but also the ability to think scientifically, understand natural phenomena logically, and apply knowledge to solve everyday problems (Bukhari et al., 2023). In a global context, science literacy has become a key indicator in various PISA international assessments (Programme for International Student Assessment). Unfortunately, the achievement of Indonesian students in PISA is still relatively low, especially in the field of science (OECD, 2024). As shown by Khurma, Jarrah, and Ali

(2025), such abilities as Intellectual curiosity and Perspective-taking has been shown to significantly affect science learning achievement, which unfortunately has not been fully integrated in science learning at the elementary level (Gillies, 2023). Therefore, there is a need for a more holistic and transformative approach to science learning from the elementary school level, so that students not only master the content, but also have thinking skills and scientific attitudes that are able to face the complexity of the rapidly changing world.

One of the factors that also affects the low science literacy of students is the quality of the evaluations used in learning, because evaluation instruments not only measure understanding, but also reflect the effectiveness of teaching methods (Barus et al., 2024; Hadianty et al., 2025). Unfortunately, in many elementary schools, the development and use of learning evaluation questions is still conventional, without going through a systematic psychometric analysis process. Assessment instruments that are invalid, unreliable, and do not have good differentiation can interfere with the accuracy in mapping student learning outcomes, because the results of the evaluation do not reflect the actual abilities of students and have the potential to produce erroneous conclusions in the learning process (Bardach & Klassen, 2020; Nirmala, 2025; Rohman, 2023). This makes assessment only an administrative tool, not an instrument of learning diagnosis and a science literacy booster.

To overcome these problems, it is necessary to develop evaluations that not only measure the cognitive aspects superficially, but are also able to reveal students' scientific thinking skills, problem-solving, and scientific attitudes. The instruments developed must meet the criteria of validity, reliability, and good differentiation, and are in line with science literacy indicators as formulated in international assessments such as PISA and TIMSS (Istyadji & Sauqina, 2023; Syach et al., 2023; Wilsa et al., 2023). In this context, it is important to design evaluations that emphasize higher-level thinking skills, as emphasized by Loyens et al (2023) that national education must be able to produce students who are creative, think critically, and be able to work together in solving increasingly complex problems (Loyens et al., 2023).

Learning evaluation instruments that are designed validly and reliably play an important role in ensuring the accuracy of measuring student learning outcomes and the effectiveness of the learning process. Instruments prepared based on psychometric principles, especially in the context of science learning, are closely related to students' metacognitive abilities and contribute significantly to the achievement of learning outcomes (Marantika, 2021; Yulianti & Herpratiwi, 2024). In the context of science

learning in elementary school, evaluation not only aims to assess mastery of content, but also measures students' scientific thinking skills, problem-solving skills, and scientific attitudes (Karengga et al., 2025). However, the reality in the field shows that the development and use of evaluation instruments in many educational units is still conventional and not yet based on psychometric principles. As a result, assessments do not reflect the students' abilities as a whole, and can even lead to erroneous conclusions in learning (Sandrika et al., 2025). Therefore, the development of valid, reliable, and diagnostic evaluation instruments is an urgent need in efforts to improve the quality of students' science literacy from the elementary school level.

This condition is even more concerning when it is associated with the results of international assessments such as the Program for International Student Assessment (PISA) which consistently shows the low science literacy of Indonesian students compared to other countries. Based on the 2022 PISA report, only 34% of Indonesian students have achieved a minimum proficiency level (Level 2) in science literacy, far below the OECD average of 76%. In fact, almost no Indonesian students fall into the high-performing category (Level 5 or 6), while the OECD average reaches 7% (OECD, 2023). Science literacy in PISA includes the ability to explain scientific phenomena, design investigations, and interpret data in real-life contexts. This fact shows that there is a real gap between the evaluation instruments used in elementary schools and the global competency indicators that are the benchmark of PISA. Therefore, it is very important to examine whether question instruments developed by paying attention to the principles of validity, reliability, differentiation, and their relationship with the context of daily life can contribute positively to improving students' science literacy (Dockterman et al., 2021).

This study aligns with recent research highlighting the critical role of assessment instrument quality in shaping science literacy outcomes among elementary school students. Hadianty et al. (2025) and Zayrin et al (2025) emphasized that invalid and unreliable test instruments hinder accurate mapping of learning outcomes. Wulandari et al. (2022) found that test items that fail to assess higher-order thinking skills tend to focus only on rote memorization rather than deep conceptual understanding. Similarly, Fitriyah & Maulida (2023) argued that the success of literacy-based assessments depends heavily on the integration of real-world contexts and scientific reasoning in test items. Furthermore, studies by Putra & Kurniawati (2021) and Rahayu et al. (2020) demonstrated that constructing items aligned with PISA and AKMI indicators is positively correlated with students' science literacy performance.

Building on these findings, this study specifically examines the psychometric evaluation of teacher-developed science test items and their implications for students' scientific literacy. The analysis focuses on the content validity, reliability, item discrimination, and alignment with international literacy indicators such as those from PISA. Although many items meet basic curriculum and technical standards, they do not necessarily contribute to improved literacy outcomes if scientific reasoning and global contexts are not integrated. Therefore, a more comprehensive psychometric evaluation is needed to ensure that assessment instruments truly support the development of students' science literacy. In this regard, the study adopts the theme Psychometric Analysis of Science Test Items and Their Implications for Scientific Literacy in Madrasah Ibtidaiyah.

METHODS

This study employed a quantitative evaluative descriptive approach aimed at comparing the quality of science test items created by Madrasah Ibtidaiyah (MI) teachers with students scientific literacy outcomes (Farid Wajdi et al., 2024). The study did not focus on developing or piloting new test items but rather analyzed existing teacher-made test documents that had been used in daily classroom assessments. The main objective was to evaluate the extent to which the quality of teacher-developed assessment instruments correlated with students' factual science literacy as reflected in their performance on Indonesian Madrasah Competency Assessment (AKMI). This approach provided an authentic evaluation of assessment practices which already implemented in real classroom settings.

The research subjects were several MI schools in a district of North Sumatra Province that had participated in the AKMI program. These schools were selected purposively because they had both AKMI performance data and teacher-developed science test items created collaboratively through the local Teacher Working Group (KKG). The primary data consisted of 30 multiple-choice science questions used in final semester exams or daily assessments. The number 30 was chosen because it represents a standard quantity for summative assessment in schools and allows for a sufficient variety of curriculum indicators and science content domains to be analyzed. This number also enables a detailed evaluation of each item without sacrificing depth or clarity in the analysis.

The analysis was conducted qualitatively and descriptively, focusing on how well each question reflected aspects of science literacy. The indicators assessed included: (1) the ability to explain scientific phenomena, (2) scientific reasoning, and (3)

the application of scientific knowledge to real-life situations. The analytical framework referred to the scientific literacy indicators developed by the Programmer for International Student Assessment (PISA) and AKMI. No statistical software such as SPSS was used in this study. Inferential statistical tests such as Pearson correlation were not applied, as the aim was not to test causal relationships but to explore patterns and tendencies between item quality and student outcomes.

The results of the teacher-made test item analysis were then compared descriptively with students' science literacy scores based on AKMI results. The comparison was presented in the form of narrative descriptions and tables to illustrate whether higher-quality items tended to align with better literacy performance. This approach enabled the researcher to objectively and contextually portray the relationship between classroom assessment quality and scientific literacy achievements among MI students

RESULTS AND DISCUSSION

This study revealed the main findings related to the quality of science questions at Madrasah Ibtidaiyah and its relationship with students' science literacy achievements. The analysis focused on three aspects: the technical construction of the questions, the integration of the indicators of high-level thinking skills (HOTS), and science literacy, as well as the tendency of the relationship between the characteristics of the questions and the science literacy score based on the AKMI assessment. Preliminary results showed that although most of the questions have met the technical standards of assessment, the coverage of the whole aspect of science literacy still needs to be improved systematically and contextually.

1. Quality of Science Questions in School Evaluation

The results of the analysis of 30 items of multiple-choice science questions made by grade V teachers of Madrasah Ibtidaiyah through KKG in Dairi Regency showed that most of the questions have met the basic criteria in the development of good evaluation instruments. The questions are compiled based on indicators that refer to the Basic Competencies in the Independent Curriculum. The preparation also considers variations in the level of difficulty of the questions, namely easy, medium, and difficult, which are distributed proportionally according to the learning evaluation guidelines. Thus, the questions reflect the balance between the demands of the curriculum and the characteristics of the students. This is in line with the findings of Juwairiah (2021), who stated that the suitability of the lesson plan and the implementation of learning based on

scientific approaches in the 2013 Curriculum are relatively good, with attention to structure, learning stages, and adjustments to competency indicators (Juwairiah, 2021).

In terms of content validity, the question items have gone through a review process by teachers and science subject coordinators in their respective madrasas. This process ensures that the questions are relevant to the curriculum and do not deviate from the learning objectives. The language used is also communicative, easy for students to understand, and free from cultural and contextual biases. The validity of the content is the key in maintaining the suitability between the teaching material and the evaluation instrument. This is in line with the findings of Yulita (2018) and Ajogbeje (2023), which tell that the development of science learning tools emphasizing the importance of content suitability with curriculum indicators, language clarity, and material readability by students as part of instrument validation by experts (Yulita, 2018), and emphasizing that the validity of the content is the most important aspect in the development of teacher-made tests, because it determines how well the question items represent the whole material to be measured substantively and factually (Ajogbeje, 2023).

The difference in the analyzed questions showed quite good performance, with the discrimination index value ranging from 0.25 to 0.40. This range of values indicates that the questions are able to distinguish students with high and low abilities effectively. Differentiation is an important indicator that shows the function of evaluation as a learning diagnostic tool. Questions with good differentiation are able to provide accurate information about student learning achievements. Therefore, this component is very important in the preparation of quality questions. This is in line with the study of Saputri et al. (2023), which emphasized that differentiation plays a role in improving the quality of questions and determining the ability of instruments to distinguish students who have or have not understood the material (Saputri et al., 2023).

The effectiveness of the distraction (incorrect answer option) also showed functional performance in most questions. Good distractions are chosen proportionally by students, not randomly selected. This signifies that students consider each answer option with understanding, not just guessing. Thus, the questions prepared not only measure memorization, but also test concept understanding. This is in line with the findings of Adawiyah and Wisudawati (2017), who emphasized that the effectiveness of distractions is important to identify functional deceptive options in measuring students' science literacy more accurately (Adawiyah & Wisudawati, 2017).

Several question items have integrated the Higher Order Thinking Skills (HOTS), especially in the application of science concepts to daily life situations. This kind of question encourages students to think critically and analytically in answering questions. This is an early indication of a shift from memorization-based assessments to competency-based assessments. This is in line with the results of Hidayati and Saparudin's (2024) study, which showed that the integration of HOTS in science education allows students to develop critical and reflective thinking in facing real-life challenges (Hidayati & Saparudin, 2024). Therefore, the existence of HOTS questions in the evaluation is a positive step that needs to be developed.

However, the overall aspect of science literacy has not been optimally integrated in the questions analyzed. Analysis showed dominance in content domains, such as factual and conceptual knowledge. Meanwhile, the domain of scientific reasoning and the applicable context is still very limited. This condition reflects that the substance of broader science literacy has not been fully reflected in the problem. Therefore, there needs to be a strengthening of the science literacy aspect in the evaluation design in the future.

The following diagram summarizes the quality performance of science questions based on seven main aspects: content validity, level of difficulty, differentiation, effectiveness of the distractor, integration of HOTS, the coverage of content domains, scientific processes, and applicable contexts in science literacy

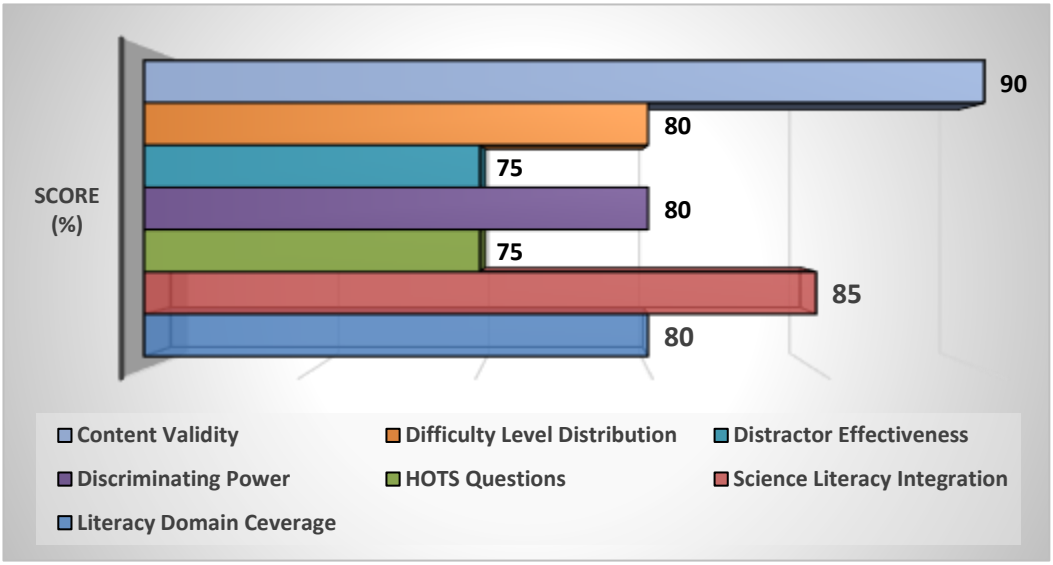


Figure 1. Bar Diagram of Quality of Science Questions Based on Psychometric Aspects and Scientific Literacy

2. Comparison of Students' Science Literacy Achievement with School Evaluation

The results of the comparison between the evaluation value of science learning in schools and the achievement of students' science literacy based on AKMI showed a significant discrepancy. A number of students who get high scores in school exams turn out to have a moderate or low AKMI score. On the other hand, it was also found that students who did not stand out in internal academic assessments actually obtained relatively high science literacy scores. This shows that conventional assessments do not fully reflect students' abilities in the aspects of scientific thinking, data evaluation, and application of knowledge in real contexts. Amalia (2023) notes that in 2021 most MI students were at the basic level in science literacy (51%), while in 2022 there was a jump to the level (59%), which indicates an increase in the ability to identify and distinguish scientific questions, although this ability has not been fully captured by school assessments (Amalia, 2023).

The following bar chart presents two important information that complement each other in the context of evaluating science learning at Madrasah Ibtidaiyah. First, the average score of AKMI (Indonesian Madrasah Competency Assessment) for the science literacy aspect in each madrasah is displayed. Second, the visualization shows the number of students in each madrasah who obtained a school science score below 70, but instead recorded an AKMI science literacy score above 65. This pattern shows a stark difference between the achievement of school-based assessments and the results of more standardized literacy assessments. These findings reinforce the indication that conventional evaluation questions in schools still measure aspects of memorization or surface understanding, rather than the ability to think scientifically or apply concepts in context. Thus, this diagram is a strong visual evidence of the importance of integrating science literacy assessments into learning evaluation practices in madrasas.

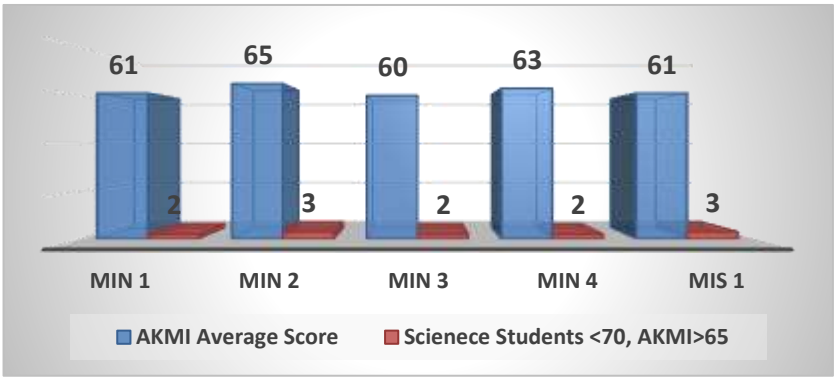


Figure 2. Comparison of AKMI Scores and Science/Science Score Mismatches

However, the overall aspect of science literacy has not been optimally integrated in the questions analyzed. Questions are still dominated by content domains, such as factual and conceptual knowledge. Meanwhile, the domain of scientific reasoning and applicative context still rarely appears. This inconsistency can be explained by the characteristics of school questions that tend to measure the ability to remember and understand concepts directly, rather than assessing the scientific thinking process. In the context of science literacy, skills such as reasoning, evaluating information, and applying knowledge in authentic situations greatly determine student achievement (Hasil PISA 2018 Volume I OECD, 2018).

The analysis showed that some students had unexpected science literacy skills. especially in the aspects of data interpretation and evidence-based reasoning. This reinforces the view that success in science literacy is not always determined by academic ability that is low in cognitive nature (Siregar et al., 2020; Syahla, 2024). Some students seem to have learning experiences that support scientific thinking skills, even if they are not reflected in school test scores. This inconsistency can be explained by the characteristics of school questions that tend to measure the ability to remember and understand concepts directly, rather than assessing the scientific thinking process (Jannah et al., 2024). Therefore, literacy-based evaluation is important as a complement in assessing students' scientific potential and development more fully, because in the context of science literacy, skills such as reasoning, evaluating information, and applying knowledge in authentic situations greatly determine student achievement.

Synchronization between the evaluation of learning at madrasas and the science literacy approach in the national assessment is an urgent need. Relevant evaluations should be able to reflect high-level thinking skills, not just content mastery (Adriantoni et al., 2025; Baderan, 2018; Yani et al., 2025, 2025). Therefore, teachers need to be equipped with the ability to formulate questions that are not only technically valid, but also measure scientific and applicative skills. This adjustment is in line with the spirit of the Independent Curriculum which emphasizes competence and contextualization. Furthermore, students' readiness to face national and global assessments will be stronger if evaluations at schools are able to integrate the principles of science literacy.

3. Implications of Findings

The findings of this study offer significant implications for science assessment practices at Madrasah Ibtidaiyah. Although most teacher-developed items meet basic technical criteria such as content validity and proportional difficulty, there remains a

notable inconsistency between item quality and students scientific literacy outcomes based on AKMI results. This suggests that existing classroom assessments are not yet optimal in measuring students' scientific reasoning abilities. This aligns with the studies of Wulandari et al. (2022) and Fitriyah & Maulida (2023), which found that context-free, recall-based questions tend to overlook essential competencies in scientific literacy.

Therefore, a reformulation of assessment design is required, emphasizing more contextual, application-oriented items aligned with global literacy indicators. Assessment should focus on evaluating students thinking processes such as data interpretation, scientific reasoning, and problem-solving rather than mere factual recall. This is supported by Rahayu et al. (2020), who advocated for a shift from rote-based testing to competency-based, and literacy-oriented assessment models.

The gap between school-based evaluations and AKMI results also highlights the need to integrate literacy-based assessment approaches into classroom practice. Teachers should be provided with training and mentoring to develop contextual items aligned with the national curriculum and international frameworks like PISA (OECD, 2024) and AKMI (Hasanah et al., 2021). Such support is essential for teachers to transition from assessment executors to assessment designers who can stimulate students' scientific thinking.

Moreover, this study underscores the urgency of implementing data-driven formative assessment in science education. Well-designed formative assessment provides accurate insights into students scientific literacy development (AlAli & Al-Barakat, 2025). Teachers can use assessment data to make informed instructional decisions and offer constructive, diagnostic feedback. In this context, teachers play a critical role as learning analysts who actively contribute to students cognitive and scientific growth (Widiyaningsih & Narimo, 2023). Thus, the study’s findings highlight not only technical concerns in item construction but also the need for a broader transformation in assessment approaches to align with literacy-focused and student-centered educational goals. …

CONCLUSION

1. Main Finding

This study found that, from a technical perspective, science test items developed by Madrasah Ibtidaiyah teachers met several basic standards of assessment quality, including content validity, proportional difficulty levels, and acceptable item

discrimination. Some items also integrated elements of Higher Order Thinking Skills (HOTS). However, these technical qualities did not align with students' actual scientific literacy performance in national assessments such as AKMI. A notable discrepancy emerged in which students who scored poorly on school tests attained high scientific literacy scores in AKMI. This phenomenon reveals that rote-based or factual assessments are insufficient to accurately measure students' scientific reasoning and real-world application skills. These findings challenge traditional assessment practices commonly used at madrasah settings and open a new discourse on the urgent need for more holistic and literacy-based evaluation designs

2. Scientific Contribution

This study contributes both theoretically and practically to the field of science education assessment at Islamic primary schools (Madrasah Ibtidaiyah). Theoretically, it supports previous research that emphasizes the inadequacy of technical item quality alone in ensuring improved scientific literacy (Zubaidah, 2019; OECD, 2023). The study also expands the scholarly discussion by highlighting the importance of process-based and contextual assessments that evaluate not only learning outcomes but also students' scientific thinking processes. Furthermore, it introduces a literacy-oriented evaluation approach that, although not relies on inferential statistics, effectively reveals discrepancies between classroom assessments and national literacy performance. In this way, the study enriches academic discourse on integrating national curricula, global literacy benchmarks such as PISA, and the local context of madrasah education.

3. Limitations

This study has several limitations. First, the sample was limited to a small number of madrasahs in a single district in North Sumatra, which restricts the generalizability of the findings to a national level. Second, only 30 multiple-choice items were analyzed, although they were selected based on alignment with national literacy indicators. Third, the analysis did not quantitatively consider student background variables such as gender, socioeconomic status, or learning styles, which may also influence AKMI outcomes. Fourth, the study did not employ inferential statistical analyses such as correlation or regression, as it adopted a descriptive-evaluative approach focused on the existing test documents. Future studies should involve a broader sample and incorporate more robust quantitative methods and additional variables to gain a deeper understanding of the causal relationship between teacher-generated assessments and students' scientific literacy achievements.

REFERENCES

Abdul Majid. Integrated Thematic Learning (First, Vol. 3). PT Remaja Rosdakarya. (2014).

Adawiyah, R., & Wisudawati, A. (2017). Pengembangan Instrumen Tes Berbasis Literasi Sains. Indonesian Journal of Curriculum and Educational Technology Studies, 5(2), 112–121. https://www.learntechlib.org/p/209353/

Adriantoni, Putri, D. N., Mayasril, N., & Gusman, R. (2025). Asesmen Abad 21: Menakar Kompetensi, Bukan Sekedar Nilai. Didaktik : Jurnal Ilmiah PGSD STKIP Subang, 11(02), Article 02. https://doi.org/10.36989/didaktik.v11i02.5970

Ajogbeje, O. J. (2023). Ensuring The Content Validity Of Teacher-Made Tests Through The Use Of A Table Of Specifications. The Seybold Report Journal, 18(09), 1223–1236. https://doi.org/10.5281/zenodo.8405151

AlAli, R. M., & Al-Barakat, A. A. (2025). Enhancing young children's science learning through science teachers' formative assessment practices. Frontiers in Education, 10. https://doi.org/10.3389/feduc.2025.1503088

Amalia, A. (2023). Two-Year Evaluation of AKMI (Indonesian Madrasah Competency Assessment) Implementation at Islamic Elementary School (Madrasah Ibtidaiyah). JENIUS (Journal of Education Policy and Elementary Education Issues), 4(2), 77–92. https://doi.org/10.22515/jenius.v4i2.7941

Baderan, J. K. (2018). Pengembangan Soal High Order Thinking (HOT) Melalui Pembelajaran Berbasis Masalah untuk Melatih Keterampilan Berpikir Kritis Peserta Didik Kelas VI SD. PEDAGOGIKA, 9(2), Article 2. https://doi.org/10.37411/pedagogika.v9i2.63

Bardach, L., & Klassen, R. M. (2020). Smart teachers, successful students? A systematic review of the literature on teachers' cognitive abilities and teacher effectiveness. Educational Research Review, 30, 100312. https://doi.org/10.1016/j.-edurev.2020.100312

Barus, R. A., Rusilowati, A., & Ridlo, S. (2024). Analisis Kebutuhan Pengembangan Instrumen Tes Penilaian Literasi Sains Berorientasi TIMSS Siswa SD Kelas V. Jurnal Pemikiran Dan Pengembangan Sekolah Dasar (JP2SD), 12(1), Article 1. https://doi.org/10.22219/jp2sd.v12i1.32712

Bukhari, H. B., Prayogi, S., & Muhali, M. (2023). Profil Literasi Sains Siswa Sekolah Dasar Pada Pembelajaran Ilmu Pengetahuan Alam dan Sosial. Bioscientist : Jurnal Ilmiah Biologi, 11(2), 2005–2014. https://doi.org/10.33394/bioscientist-.v11i2.14448

Dockterman, D., Petscher, Y., McAfee, A., Klopfer, E., Osterweil, S., & Diefenthaler, C. (2021). Game Design Considerations for Screening, Interim, and Diagnostic Educational Assessments. Practical Assessment, Research, and Evaluation, 26(1), Article 1. https://doi.org/10.7275/22109276

Hanik, E. U., & Dewanti, S. S. (2025). Ethnosocial Learning Based on Socio-Cultural Literacy: An Exploratory Study in Elementary School. Journal of Integrated Elementary Education, 5(1), 254-268.

Farid Wajdi, Rokhamah Rokhamah, Fitri Puji Lestari, Maemunah Maemunah, Hery Yanto The, Mardiati Mardiati, Yeni Rahmawati, Mardiana Mardiana, Irwanto Irwanto, Heny Ekawati Haryono, Inten Nurmalasari, Rory Anugraha, & Hapizil Umam. (2024). Metodologi Pendidikan (Teori Dan Aplikasi Dalam Penelitian) (Evi

Damayanti, Ed.). CV WIDINA MEDIA UTAMA.

Febriandani, R., Yetti, E., & Utami, A. D. (2025). Exploring Science Literacy for Early Childhood: Eksplorasi Literasi Sains untuk Anak Usia Dini. PAUDIA: Jurnal Penelitian Dalam Bidang Pendidikan Anak Usia Dini, 24–37. https://doi.org/10.26877/paudia.v14i1.1118

Gillies, R. M. (2023). Dialogic teaching in a Year 5 classroom during cooperative inquiry-based science. International Journal of Educational Research Open, 5, 100290. https://doi.org/10.1016/j.ijedro.2023.100290

Hadianty, H., Rochman, C., & Rahmawati, R. (2025). Evaluasi Hasil Belajar Siswa Pada Materi Bioteknologi dalam Meningkatkan Kemampuan Literasi Sains: (Evaluation of Student Learning Outcomes on Biotechnology Material in Improving Scientific Literacy Skills). BIODIK, 11(1), Article 1. https://doi.org/10.22437/biodik.-v11i1.41244

Hanik, E. U., & Laili, N. F. (2024, August). Implementation Of Differentiation Learning With The Project Based Learning Model In Primary School IPAS Learning. In ICIE: International Conference on Islamic Education (Vol. 4, pp. 245-258).

Hasanah, U., Edwita, & Januar, A. (2021). Pendampingan Guru Mengembangkan Assesment Kompetensi Minimum (Akm) Berorientasi Pisa Untuk Meningkatkan Kualitas Hasil Pembelajaran Di Sekolah Dasar Wilayah Kabupaten Bogor. Jurnal Abadimas Adi Buana, 5(01), Article 01. https://doi.org/10.36456/abadimas.-v5.i01.a3634

Hasil PISA 2018 Volume I OECD. (2018). https://www.oecd.org/en/publications/pisa-2018-results-volume-i_5f07c754-en.html

Hidayati, B. T., & Saparudin, S. (2024). Higher Order Thingking Skills (Hots) dalam Konsep Integrasi Pendidikan Islam dan Sains. EduInovasi: Journal of Basic Educational Studies, 4(1), Article 1. https://doi.org/10.47467/edui.v4i1.5800

Istyadji, M., & Sauqina. (2023). Conception of scientific literacy in the development of scientific literacy assessment tools: A systematic theoretical review. Journal of Turkish Science Education, 20(2), Article 2. https://doi.org/10.36681/tused.-2023.016

Jannah, R., Khairunnisa, Sholeh, M., & Khaliq, R. (2024). Minimum Competency Assessment in Madrasah Ibtidayah. Al-Hayat: Journal of Islamic Education, 8(3), Article 3. https://doi.org/10.35723/ajie.v8i3.573

Juwairiah, J. (2021). Evaluasi Penggunaan Pendekatan Saintifik di dalam Pembelajaran Siswa Sekolah Dasar. Journal of Integrated Elementary Education, 1(2), Article 2. https://doi.org/10.21580/jieed.v1i2.7717

Karengga, F. I., Rizko, U., & Bashith, A. (2025). Analisis Problematika Pelaksanaan Evaluasi Pembelajaran IPA dalam Mencapai Tujuan Pendidikan pada Kurikulum Merdeka SD/MI. Al-Madrasah: Jurnal Ilmiah Pendidikan Madrasah Ibtidaiyah, 9(2), Article 2. https://doi.org/10.35931/am.v9i2.4401

Latifah, F., & Hanik, E. U. (2023). Penerapan Model Problem Based Learning Menggunakan Pendekatan STEM Untuk Meningkatkan Keterampilan Pemecahan Masalah Siswa Kelas IV Mata Pelajaran IPA MI An-Nur Daren. Jurnal Jendela Pendidikan, 3(02), 171-179.

Loyens, S. M. M., van Meerten, J. E., Schaap, L., & Wijnia, L. (2023). Situating Higher-Order, Critical, and Critical-Analytic Thinking in Problem- and Project-Based

Learning Environments: A Systematic Review. Educational Psychology Review, 35(2), 39. https://doi.org/10.1007/s10648-023-09757-x

Marantika, J. E. R. (2021). Metacognitive ability and autonomous learning strategy in improving learning outcomes. Journal of Education and Learning (EduLearn), 15(1), Article 1. https://doi.org/10.11591/edulearn.v15i1.17392

Nirmala, M. A. (2025). Pengaruh Metode Mind Mapping Terhadap Hasil Belajar PKN Siswa Kelas IV di SD Negeri 1 Negara Bumi Udik [Undergraduate, IAIN Metro]. https://repository.metrouniv.ac.id/id/eprint/11297/

OECD. (2023). PISA 2022 Results (Volume I): The State of Learning and Equity in Education. OECD. https://doi.org/10.1787/53f23881-en

OECD. (2024). PISA 2022 Results (Volume III): Creative Minds, Creative Schools. OECD. https://doi.org/10.1787/765ee8c2-en

Rohman, S. (2023). Pengembangan Instrumen Penilaian Hasil Belajar Berbasis Taksonomi Bloom. Al-Ibrah : Jurnal Pendidikan dan Keilmuan Islam, 8(1), Article 1. https://doi.org/10.61815/alibrah.v8i1.262

Sandrika, T., Kartika, T. A., Hasibuan, T. K., Akil, A., & Azis, A. (2025). Transformasi Penilaian Pembelajaran di Kelas dalam Pembelajaran Holistik di Era Kurikulum Merdeka. Hayati : Journal of Education, 1(1), Article 1. https://jurnal.si-nesia.id/index.php/hayati/article/view/344

Saputri, H. A. S., Zulhijrah, Larasati, N. J., & Shaleh. (2023). Analisis Instrumen Assesmen: Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Beda Butir Soal. Didaktik : Jurnal Ilmiah PGSD STKIP Subang, 9(5), Article 5. https://doi.org/-10.36989/didaktik.v9i5.2268

Siregar, T. R. A., Iskandar, W., & Rokhimawan, M. A. (2020). Literasi Sains melalui Pendekatan Saintifik pada Pembelajaran IPA SD/MI di Abad 21. MODELING: Jurnal Program Studi PGMI, 7(2), Article 2. https://doi.org/-10.69896/modeling.v7i2.582

Syach, A., Sugandi, D., & Yusup, S. H. (2023). Penerapan Metode Eksperimen untuk Meningkatkan Hasil Belajar Siswa dalam Mata Pelajaran IPA Pada Sub Pokok Bahasan Sifat-Sifat Cahaya. Jurnal Primary Edu, 1(1), Article 1. https://jurnal.rakeyansantang.ac.id/index.php/primary/article/view/388

Syahla, A. F. (2024). Pengembangan E-Book Berbasis Socioscientific Issue untuk Meningkatkan Kemampuan Berpikir Kritis Siswa [Diploma, UIN Raden Intan Lampung]. https://repository.radenintan.ac.id/32095/

Trisnawati, W., Musanada, N., & Hanik, E. U. (2022). Peningkatan Hasil Belajar Peserta Didik dengan Menerapkan Metode Problem Based Learning (PBL) melalui Media Gambar pada Pelajaran IPA Kelas V SDN 05 Ngembalrejo Bae Kudus. PANDAWA, 4(1), 154-165

Widiyaningsih, P., & Narimo, S. (n.d.). Peran Guru dalam Memaksimalkan Semangat Belajar Peserta Didik pada Implementasi Program Kurikulum Merdeka di SMK Negeri 1 Boyolali | JIIP - Jurnal Ilmiah Ilmu Pendidikan. Retrieved July 24, 2025, from http://jiip.stkipyapisdompu.ac.id/jiip/index.php/JIIP/article/view/2753

Widiyaningsih, P., & Narimo, S. (2023). Peran Guru dalam Memaksimalkan Semangat Belajar Peserta Didik pada Implementasi Program Kurikulum Merdeka di SMK Negeri 1 Boyolali. JIIP - Jurnal Ilmiah Ilmu Pendidikan, 6(8), 6325–6332. https://doi.org/10.54371/jiip.v6i8.2753

Wilsa, A., Rusilowati, A., Susilaningsih, E., Jaja, J., & Nurpadillah, V. (2023). Validity, reliability, and item characteristics of cell material science literacy assessment instruments. Jurnal Penelitian Dan Evaluasi Pendidikan, 27(2), 177–188. https://doi.org/10.21831/pep.v27i2.61577

Yani, I. D. A. D. A., Lestari, A. P. I. Y., Sudianta, I. N., Mahayuni, N. K. A., & Mursalin, M. (2025). Model Evaluasi Holistik Berbasis Rubrik Kombinasi Dalam Menilai Literasi Awal Siswa Sekolah Dasar. Elementary School: Jurnal Pendidikan Dan Pembelajaran Ke-SD-An, 12(2), Article 2. https://doi.org/10.31316/-esjurnal.v12i2.4527

Yulianti, D., & Herpratiwi, H. (2024). Development of a science, environment, technology, and society-based learning module to foster critical thinking in elementary students. Journal of Education and Learning (EduLearn), 18(4), Article 4. https://doi.org/10.11591/edulearn.v18i4.21713

Yulita, E. (2018). Pengembangan Modul Pembelajaran IPA Berbasis Inkuiri Terbimbing untuk Meningkatkan Keterampilan Proses Dasar Sains Peserta Didik Kelas IV MI/SD. JMIE (Journal of Madrasah Ibtidaiyah Education), 2(2), Article 2. https://e-journal.adpgmiindonesia.com/index.php/jmie/article/view/70