



Analysis of Teachers' and Students' Needs in the Development of Two-Tier Multiple-Choice Questions on Environmental Issues in Ecology and Environmental Pollution

Mutiara Putri Milzandi, Novi Ratna Dewi*

Science Education Program, Universitas Negeri Semarang, Semarang, Indonesia

✉ * noviratnadewi@mail.unnes.ac.id

Article Info	Abstract
Keywords: <i>Tapak Jedah; open dumping; protected forest; community-based waste management; Bengkulu City;</i> Received: 03/08/2026 Revised: 08/08/2026 Accepted: 13/08/2026	Background: The assessment of Higher-Order Thinking Skills (HOTS) is a crucial aspect of 21st-century science education. However, current evaluation instruments in schools are still dominated by Lower-Order Thinking Skills (LOTS) questions in the form of conventional multiple-choice questions, which trigger a “guessing effect” among students and are unable to accurately map conceptual understanding Methods: This study is a preliminary research project employing a mixed-methods approach with a sequential explanatory design. Participants consisted of 421 junior high school students and 30 science teachers selected using purposive sampling. Data collection was conducted through Likert-scale questionnaires and open-ended questions. Data analysis used descriptive percentages for quantitative data and thematic analysis aided by NVivo software for qualitative data. Results: The findings revealed that 60.83% of teachers still predominantly use LOTS-standardized assessments, while 70.33% of students often guess answers at random when encountering difficulties. The integration of real-world environmental case studies was highly favored by 81.01% of students because it helped make abstract ecological concepts more concrete. A total of 84.17% of teachers expressed an urgent need for a valid two-tier multiple-choice instrument, hoping for stimulus designs that are concise, visual, and feature a practical scoring system. Conclusion: There is a high urgency to develop a two-tier multiple-choice instrument focused on environmental issues as an objective, diagnostic, and practical solution for evaluating Higher-Order Thinking Skills (HOTS).

INTRODUCTION

Educational transformation in Indonesia aims to develop students' potential through contextual, meaningful, and challenging learning processes (Hunaepi & Suharta, 2024; Thaariq, 2026). One key aspect emphasized is strengthening higher-order thinking skills (HOTS). Higher-

order thinking skills are among the essential skills students must possess in learning during the era of the Fourth Industrial Revolution (Komalasari et al., 2023). The importance of higher-order thinking skills in learning has become a primary focus, marking a shift in perspective toward efforts to cultivate a generation capable of adapting to advancements in science, technology, engineering, and mathematics (STEM) in the 21st century. HOTS abilities include analyzing (C4), evaluating (C5), and creating (C6), which play a crucial role in preparing students to face global challenges (T. Li et al., 2026). HOTS plays a vital role in equipping students with critical, analytical, and creative thinking skills so they are prepared to face changes and challenges in today's digital and information age (Sri Hanipah, 2023).

In the digital and information age, it is not enough for students to memorize course material. They must be able to process information, make sound decisions, and solve problems independently and logically (Cynthia & Sihotang, 2023; Mahmud & Wong, 2022). Thus, mastery of HOTS serves as a key indicator in assessing the quality of education. However, to achieve this goal, appropriate assessment tools or instruments are also needed that can accurately and objectively represent HOTS abilities. HOTS assessment instruments help train students to think critically, analyze, and solve problems independently. In addition, teachers can also obtain more accurate feedback on students' learning progress so that they can provide appropriate interventions (Adzidzah & Yudiawan, 2024; T. Li et al., 2026). According to Rapti et al. (2026), the development of HOTS is expected to help students master four key 21st-century competencies: critical thinking, creativity, communication, and collaboration (4K).

Although the importance of higher-order thinking skills has been widely publicized, in practice there are still various challenges, particularly regarding the measurement or assessment of these skills. Evaluation instruments commonly used in schools, such as conventional multiple-choice questions, still largely focus on measuring lower-order thinking skills (LOTS), such as recall (C1) and comprehension (C2) (Ayaz & Abbas, 2026). This is also consistent with the findings of Riza Andriani et al. (2024), who reported that teachers' understanding of Bloom's taxonomy as the foundation for developing HOTS remains low: only 13.33% of teachers have a strong understanding, while 53.33% have only ever attempted to create HOTS questions. The majority (75%) consider the process difficult to very difficult. These limitations can affect teachers' ability to translate higher-order cognitive demands into test item construction. Developing a HOTS assessment cannot be done simply by identifying operational verbs that represent higher-order thinking skills. Determining whether a task qualifies as HOTS requires consideration of the thinking skills construct to be measured, as well as the characteristics of the task that can elicit higher-order cognitive processes (Lian, 2023; Zhou et al., 2023). When these competencies have not yet been fully mastered, questions that are nominally categorized as HOTS questions may still lead students to simply recall or understand information. As a result, students' abilities to analyze problems, evaluate information, and create solutions have not been fully captured through the current assessment processes. Therefore, there is a critical need to develop innovative and structured assessment instruments capable of measuring higher-order thinking skills in a valid, reliable, and contextually appropriate manner for today's learning environment.

One alternative solution to measure higher-order thinking skills more thoroughly is a two-tier multiple-choice test. The two-tier multiple-choice test was developed by Treagust (1988) at Curtin University in Australia. This test consists of two parts: the first part consists of multiple-choice questions, and the second part asks students to provide a rationale for the answer they selected in the first part (Sibiç et al., 2022). Research conducted by Hidayatullah et al. (2022) on the development of a two-tier multiple-choice test based on higher-order thinking skills (HOTS) for the topic of Acids, Bases, and Salts showed that the use of HOTS questions in schools is highly effective for improving students' higher-order thinking skills through the two-tier multiple-choice test. Thus,

this instrument is highly relevant for measuring higher-order thinking skills, as it encourages students not only to select an answer but also to justify their reasoning logically (Andriyatno et al., 2023). Furthermore, two-tier tests provide richer diagnostic information for teachers, enabling them to design more targeted follow-up learning strategies (He et al., 2022). The literature cited above indicates that these characteristics can result in assessments that differ from single-level, multiple-choice tests. One study comparing the two formats, conducted by Samaduri (2022), showed that students' level of understanding on standard multiple-choice tests is often difficult to describe, as the answers are limited to a few predetermined options. In contrast, in reasoned multiple-choice tests, students' level of understanding can be assessed. This finding is important because it shows that adding a second level not only makes the questions more difficult but also enhances the instrument's ability to distinguish students' levels of competency.

Amid today's various global challenges, environmental issues have become one of the most pressing concerns requiring attention and concrete action from all segments of society, including the education sector (Molnár et al., 2026; Uralovich et al., 2023). In this context, students, as the next generation, play a crucial role in preserving the environment through sustainable understanding and action. Integrating environmental issues into contextual learning has proven effective in connecting scientific theory with real-world realities, while also enhancing ecological awareness (Fajri et al., 2026; Tahmid et al., 2024). Furthermore, using environmental issues as an assessment context not only instills an environmentally conscious character but is also highly relevant for developing students' abilities to communicate, think critically, and solve problems (Purwanto et al., 2022). Thus, integrating environmental issues into assessment instruments not only improves the quality of measuring students' cognitive abilities but also supports the achievement of sustainable education goals.

Although the potential of a two-tier multiple-choice test incorporating environmental issues is immense, developing this instrument cannot be done instantly without a strong data foundation from the target users. The success of an assessment tool heavily depends on its alignment with students' characteristics and the real challenges faced by teachers in the classroom (Ketterlin-Geller et al., 2024). Without a clear needs assessment, the developed instrument risks being impractical or too difficult for junior high school students. In line with this, Dewi et al. (2025) emphasize the importance of developing validated assessment instruments to improve the quality of science learning evaluation. Thus, before proceeding to the product development stage, an in-depth study of the objective conditions in schools must be conducted. This study aims to analyze the needs of teachers and students in developing two-tier multiple-choice test items addressing environmental issues in the topics of ecology and environmental pollution. The results of this analysis are expected to serve as a fundamental basis for designing valid, practical, and relevant HOTS assessment instruments that meet the needs of contemporary science education.

METHODS

This study is preliminary research within the Research and Development (R&D) framework using the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) model, employing a mixed-methods approach with an explanatory sequential design. The objective is to analyze the needs of students and teachers regarding the development of two-tier multiple-choice instruments incorporating environmental issues to measure Higher Order Thinking Skills (HOTS). The present study specifically focuses on the Analysis stage of the ADDIE model. The subsequent Design, Development, Implementation, and Evaluation stages represent the planned stages for the subsequent development and refinement of the assessment instrument. The research was conducted during the even semester of the 2025/2026 academic year. Selected research subjects

were through purposive sampling and participated voluntarily, including 421 junior high school students who had studied Ecology and Environmental Pollution, as well as 30 science teachers.

Data collection was conducted simultaneously using a questionnaire containing two types of questions. First, a closed-ended questionnaire with a 4-point Likert scale was used to quantitatively measure experiences with existing assessments, environmental literacy, and the urgency of a two-tier instrument. Second, open-ended questions were used to gather qualitative data regarding specific difficulties with the subject matter (for students) as well as challenges in developing Higher-Order Thinking Skills (HOTS) questions and expectations for the instrument's design (for teachers). The data collected were analyzed using two analytical techniques tailored to the characteristics of the data:

Quantitative Data Analysis

To analyze data from closed-ended Likert-scale questionnaires, descriptive statistical techniques involving percentages were used. The response scores of all respondents were accumulated, and the percentages were calculated using the following formula:

$$P = \frac{f}{N} \times 100\%$$

Explanation:

P = Percentage of suitability/need (%)

f = Total score obtained from all respondents

N = Maximum total score

Qualitative Data Analysis

To analyze data from open-ended questions, we used a thematic analysis approach aided by CAQDAS (Computer-Assisted Qualitative Data Analysis Software), specifically NVivo. The stages of qualitative data analysis include:

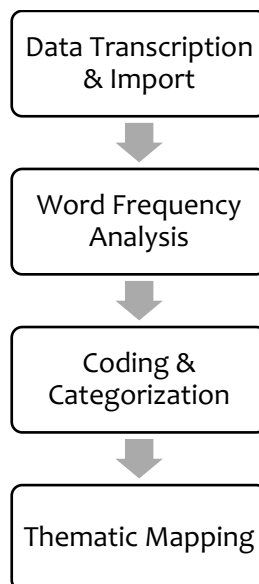


Figure 1. Stages of qualitative data analysis

RESULT AND DISCUSSION

In the 21st century, science education evaluation should ideally focus on a comprehensive assessment of higher-order cognitive abilities, rather than merely memorizing facts. However, field observations reveal a discrepancy between teachers' pedagogical perceptions and the

implementation of assessment in the classroom. This discrepancy in implementation is presented in Table 1 below.

Table 1. Aspects of Assessment and Measurement Implementation According to Teachers

Statement	Results
I have been regularly developing and using assessment tools to measure students' higher-order thinking skills (HOTS)	78.33%
In my opinion, HOTS-type questions are very important to implement in order to develop students' analytical skills	93.33%
I rarely assign HOTS-standard questions because students often have difficulty understanding them.	55.83%
I fully understand how to develop question indicators to measure students' evaluation and creative skills	74.17%
The assessments I currently use measure memorization and comprehension (LOTS) more than higher-order analytical skills	60.83%

Source: Researcher's Document

Nearly all teachers (93.33%) agree that HOTS assessment tools are very important for helping students improve their analytical skills. However, in practice, 60.83% of teachers agree that the assessment instruments used are still dominated by items in the low-order thinking skills (LOTS) domain, particularly at the cognitive levels of recall (C1) and comprehension (C2). Consequently, students' cognitive patterns when answering exam questions are directly influenced by the dominance of these conventional instruments.

Table 2. Aspects of Students' Experiences in Answering Questions

Statement	Results
So far, science questions have more often used standard multiple-choice questions (selecting only A, B, C, or D)	80.52%
If I don't know the answer to a multiple-choice question, I often just guess the answer	70.33%
Sometimes my answer is correct on a standard multiple-choice question, even though I don't actually know the reason behind that answer.	69.42%

Source: Researcher's Document

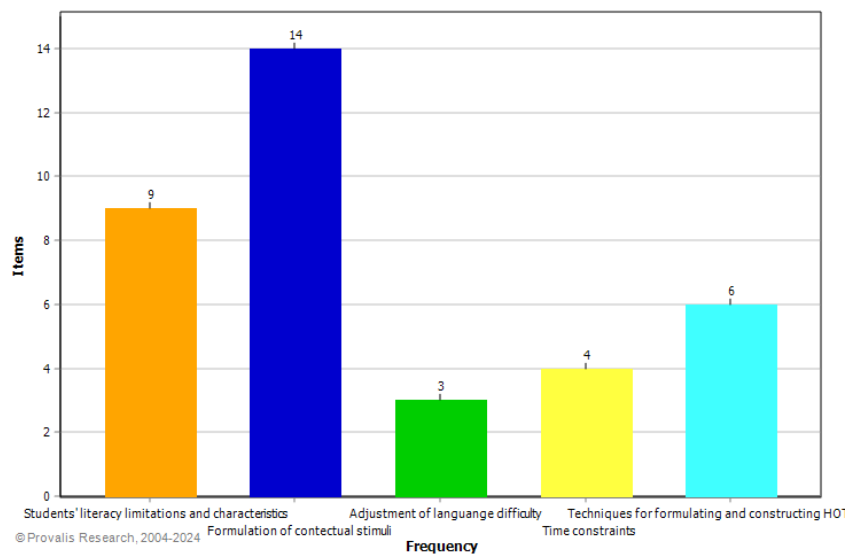
Table 2 illustrates students' experiences with science questions. A total of 80.52% of students stated that they are already accustomed to the standard multiple-choice format. When students face exam questions with a high level of difficulty, this triggers poor academic behavior: 70.33% of students stated that they often guess the answers. In addition, 69.42% of students stated that their answers were sometimes correct by chance, rather than due to a deep understanding of the material. This confirms the occurrence of a mass "guessing effect" in science assessment. Research (Jin et al., 2022) indicates that this phenomenon can increase error variance and result in an overestimation of students' abilities. Consequently, students' final scores fail to represent their actual cognitive structure or mastery of the material but are instead largely influenced by sheer luck. The sustained use of the LOTS instrument in schools is hindered by various operational and technical challenges faced by teachers. There is a gap between teachers' conceptual understanding of this innovative instrument and its practical application in the classroom.

Table 3. Aspects of Teachers' Understanding and Use of the Two-Tier Multiple-Choice Instrument

Statement	Results
I understand the concept of the <i>Two-Tier Multiple Choice</i> assessment instrument	76.67%
I have never used the <i>Two-Tier Multiple Choice</i> question format to assess students' understanding	65.83%
The use of the <i>Two-Tier</i> instrument is very effective for identifying students' misconceptions regarding environmental pollution material	80.83%
Creating <i>Two-Tier multiple-choice</i> questions takes too much time, so I am reluctant to use them in class.	58.33%
I need a variety of assessment formats beyond standard multiple-choice questions to measure students' thought processes and deep conceptual understanding	90.83%

Source: Researcher's Document

Based on the data in Table 3, we found inconsistencies in implementation. Most teachers (76.67%) understand the Two-Tier Multiple-Choice instrument conceptually, and 80.83% consider it useful for identifying students' misconceptions about environmental pollution. However, 65.83% of teachers stated that they have never assessed student progress using this instrument. A total of 58.33% of teachers stated that developing Two-Tier Multiple-Choice questions takes too much time, thereby reducing their willingness to use it independently. This is the primary technical barrier limiting its implementation. However, 90.83% of teachers stated that there is a significant need for an assessment instrument capable of accurately measuring students' cognitive abilities.

**Figure 2.** Visualization of Coding Frequency for Teachers' Challenges

Based on the Coding Frequency Visualization (Figure 2), the most dominant challenge faced by teachers is "Limitations in literacy and student characteristics." Secondary obstacles identified include difficulties in "Formulating contextual stimuli," "Adjusting language proficiency levels," and "Time constraints." These findings are supported by several studies that found the main challenges teachers face when designing HOTS assessments relate to the complexity of creating contextual and relevant stimuli, as well as formulating test items regarding sentence structure and word choice (Otaya et al., 2025; Risni & Khotimah, 2025; Saepuzaman et al., 2022).

Furthermore, the literature (Riza Andriani et al., 2024) indicates that these technical limitations are also linked to teachers' still-limited understanding of Bloom's taxonomy as the foundation for developing HOTS. Although teachers acknowledge the importance of HOTS, they

required them to analyze a case study or empirical data. Thus, the structured integration of real-world environmental issues into assessment instruments is a crucial step toward addressing these difficulties.

Table 5. Aspects of Student Interest in Environmental Issues

Statement	Results
I am more interested in working on problems that use stories or real-life environmental cases	81.01%
Reading questions that address real-world environmental issues makes it easier for me to visualize the subject matter and helps me develop my critical thinking skills	83.86%
In my opinion, studying ecology would be more meaningful if it were always linked to efforts to solve current environmental problems.	79.92%

Source: Researcher's Document

Based on the data in Table 5 regarding students' interest in environmental issues, the results show that students welcome the inclusion of environmental issues in test questions. A total of 81.01% of students are more interested in working on questions based on real-world environmental problems. Additionally, 83.86% of students feel that questions addressing real-world problems help them visualize abstract concepts more clearly while also developing their critical thinking skills. In line with this, 79.92% of students felt that ecology lessons were more meaningful when directly linked to efforts to solve current environmental problems. These positive views among students align with teachers' perceptions regarding the integration of environmental issues.

Table 6. Aspects of Environmental Issue Integration

Statement	Results
I always try to relate ecological and environmental pollution topics to real-world environmental issues occurring today	88.33%
Including local or global environmental problems in exam questions can foster students' environmental awareness and concern	89.17%
I feel that the textbook is sufficient, so there is no need to add current environmental issues as evaluation material.	44.17%
Students will be more engaged and better able to think critically if questions are presented in the form of contextual environmental damage cases.	85%

Source: Researcher's Document

Based on Table 6, many teachers (88.33%) have made efforts to link learning materials to current environmental issues, and they also recognize the benefits of this approach. 89.17% of teachers agree that presenting local and global environmental problems effectively fosters students' awareness, and 85% believe that presenting cases of environmental damage effectively promotes critical thinking skills. In the learning process, using environmental issues helps foster students' abilities to communicate, think critically, and solve problems (Uralovich et al., 2023).

Based on the identification of issues related to the "guessing effect," the challenges in formulating HOTS questions, and the need for contextual integration, this study highlights the urgency of developing an instrument capable of measuring HOTS while also offering a high level of practicality for users. This need is clearly reflected in the responses of students and teachers regarding the development of a two-tier multiple-choice instrument.

Table 7. Students' Perceptions of Two-Tier Multiple-Choice Questions

Statement	Results
I have worked on questions where I had to select an answer (A/B/C/D) and also had to select the reason for my answer (two-tier)	48.91%
In my opinion, question formats that require an explanation for the answer will prevent me from the habit of guessing randomly.	80.46%
Answering questions along with my reasoning can honestly demonstrate the extent to which I truly understand the material being tested.	85.56%

Source: Researcher's Document

From the students' perspective, the level of acceptance of this new question format shows very positive results. The data in Table 7 indicates that students' initial familiarity with the two-tier format is still categorized as low (48.91%), meaning many of them are not yet accustomed to this format. Nevertheless, 80.46% of students agree that the requirement to select a justification at the second level helps reduce speculative behavior or random guessing. Additionally, 85.56% of students feel confident that answering questions along with their justifications can represent their level of conceptual mastery more objectively and authentically. This high level of awareness aligns with students' expectations regarding learning assessment, as shown in Table 8 below.

Table 8. Students' Needs and Expectations

Statement	Results
I agree that teachers should provide practice questions that not only test final results but also train my way of thinking	84.28%
I agree that reasoned multiple-choice (two-tier) questions should be developed that include real-world environmental scenarios to train my analytical skills.	78.88%
If these practice questions are ready, I am willing to work on them to reinforce my understanding of Ecology and Environmental Pollution.	77.43%

Source: Researcher's Document

Based on Table 8, 84.28% of students agreed with implementing an assessment tool that not only measures outcomes but also trains reasoning skills and thought processes. Thus, 78.88% of students fully supported the development of a two-tier assessment tool featuring real-world environmental cases, and 77.43% expressed their willingness to use the tool to consolidate their understanding of the material.

Table 9. Aspects of Teachers' Needs for the Development of New Assessment Instruments

Statement	Results
I need a validated two-tier multiple-choice assessment instrument specifically for ecology and environmental pollution topics	84.17%
Currently, there are already many available two-tier question banks covering environmental issues, so there is no need to develop a new instrument.	55.83%
The development of this assessment instruments focused on environmental issues will greatly help me evaluate students' higher-order thinking skills (HOTS) more comprehensively	83.33%
I feel that standard multiple-choice instruments (single tier) are sufficient for measuring HOTS without the need for two-tier instruments.	52.50%
I am willing to implement this two-tier multiple-choice instrument focused on environmental issues if it is successfully developed in the future	86.67%

Source: Researcher's Document

Teachers also identified similar needs. The data in Table 9 show that 84.17% of teachers expressed a fundamental need for a validated two-tier instrument, specifically on the topics of ecology and environmental pollution. In line with this, 83.33% of teachers believe that instruments incorporating environmental issues will facilitate a more comprehensive evaluation of students' HOTS abilities. Although some teachers (52.50%) still consider conventional multiple-choice instruments adequate and assume that the existing question bank is sufficient (55.83%), the goal of adopting innovative new instruments remains predominant. This is evidenced by the high percentage of teachers (86.67%) who are willing to implement a two-tier instrument incorporating environmental issues if such an instrument is developed. The design or specifications of the instrument expected by teachers are illustrated in the coding frequency visualization below.

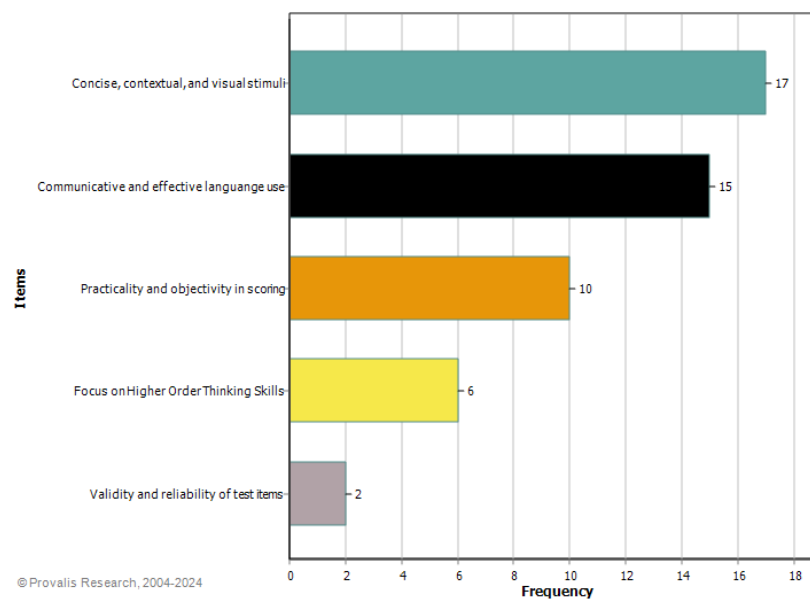


Figure 4. Visualization of Coding Frequency for Teachers' Expectations

The results of the qualitative analysis in Figure 4 reveal three main expectations teachers have for the instrument to be developed: “Brief, contextual, and visual stimuli,” “Easy-to-understand language,” and “A practical scoring system.” These needs are well-founded from a pedagogical perspective. Based on a systematic review of working memory function, processing overly complex linguistic information or text consumes massive amounts of cognitive resources, thereby degrading the remaining memory capacity needed for higher-order cognitive processes (S. Li, 2023). In the context of science assessment, presenting excessive text-based narrative stimuli or using inaccessible language structures risks triggering working memory overload. Conversely, images, infographics, or diagrams can help students process information more efficiently, allowing their cognitive focus to be directed toward the analytical process (Jadallah Abed Khasawneh & Ahmad Saleem Khasawneh, 2023; Jee et al., 2022). Thus, the two-tier multiple-choice instrument on environmental issues used in this study was specifically designed to address all of these needs namely, to provide an unbiased assessment tool for students and to offer a valid, informative, and practical HOTS evaluation instrument for teachers.

CONCLUSION

Based on the results of the needs analysis, it can be concluded that developing an innovative assessment instrument for the subject matter of Ecology and Environmental Pollution is crucial. Current assessment practices are still dominated by Lower Order Thinking Skills (LOTS)

instruments in the form of conventional multiple-choice questions, which lead to a high incidence of guessing and thus fail to represent students' cognitive abilities objectively. On the other hand, teachers face crucial challenges in independently developing Higher Order Thinking Skills (HOTS) instruments, including time constraints, difficulties in formulating contextual stimuli, and obstacles in adapting to students' language proficiency levels. These difficulties are compounded by the conceptual and abstract nature of ecology material (such as energy flows, biogeochemical cycles, and pollution dynamics). Based on the findings of this needs analysis, it is recommended that future research focus on the design and development stages of an actual product in the form of a two-tier multiple-choice instrument addressing environmental issues. Once the instrument has been successfully developed and validated by experts, an empirical trial is needed to quantitatively demonstrate the instrument's effectiveness in reducing the guessing effect compared to conventional tests, as well as to assess the practicality of its scoring system for teachers. In addition, the validated instrument can later be used by other researchers to specifically map students' misconceptions regarding ecology. If this two-tier framework for analysis and instrument development is to be adapted to different science topics, there are several limitations that must be considered. First, the formulation of contextual stimuli must be carefully adjusted, if environmental issues are not relevant, researchers need to identify appropriate socio-scientific issues (SSIs) or everyday technological phenomena so that the stimuli do not seem forced. Second, in exact-calculation material, the design of second-tier (tier 2) reasoning options demands much stricter management of cognitive load. The combination of complex mathematical formulas and narrative text analysis in two-tier questions risks causing a double overload on working memory, which could actually obscure the original purpose of the HOTS assessment itself.

DECLARATION OF GENERATIVE AI

During the preparation of this work, the authors used Grammarly and Google Gemini to assist with grammar, language refinement, and writing improvement. After using this tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CONFLICTING INTEREST STATEMENT

The author(s) declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FUNDING INFORMATION STATEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHORSHIP CONTRIBUTION STATEMENT

Mutiara Putri Milzandi: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing – Original Draft, and Visualization. **Novi Ratna Dewi:** Supervision, Validation, and Writing – Review & Editing.

REFERENCES

- Adzidzah, N., & Yudiawan, A. (2024). HOTS-based formative assessment: The key to improving the quality of learning. *Journal of Quality Assurance in Islamic Education*, 4(2), 109–120.
- Andriyatno, I., Zulfiani, Z., & Mardiaty, Y. (2023). Higher order thinking skills: Student profile using two-tier multiple choice instrument. *International Journal of STEM Education for Sustainability*, 3(1), 111–124.
- Ayaz, N., & Abbas, S. G. (2026). A content analysis of board examination papers: Mapping cognitive domain levels through Bloom's taxonomy. *Pakistan Journal of Social Sciences Review*, 2959–8044.
- Azrai, E. P., Japar, M., & Situmorang, R. (2025). Diagnosis of ecosystem misconceptions for high school students in Jakarta. *International Journal of Evaluation and Research in Education*, 14(5), 3790–3799.
- Cynthia, R. E., & Sihotang, H. (2023). Melangkah bersama di era digital: Pentingnya literasi digital untuk meningkatkan kemampuan berpikir kritis dan kemampuan pemecahan masalah peserta didik. *Jurnal Pendidikan Tambusai*, 7(3), 31712–31723.
- Dewi, N. R., Listiaji, P., & Savitri, E. N. (2025). Development and validation of an ICT-based TPACK test for pre-service science teachers: Toward objective and scalable assessment. *Journal of Physics: Conference Series*, 3148(1), 012014.
- Fajri, N., Arnyana, I. B. P., Adnyana, P. B., Citrawathi, D. M., Fajri, S. R., Sarnita, F., & Fitriani, H. (2026). Integration of local potential and innovative learning methods to improve environmental awareness: A systematic literature review. *Journal of Turkish Science Education*, 23(2), 417–435.
- He, P., Zheng, C., & Li, T. (2022). Upper secondary school students' conceptions of chemical equilibrium in aqueous solutions: Development and validation of a two-tier diagnostic instrument. *Journal of Baltic Science Education*, 21(3), 428–444.
- Hidayatullah, A. R., Yamtinah, S., & Masykuri, M. (2022). Development of a two-tier multiple-choice instrument based on higher order thinking skills (HOTS) on acids, bases, and salts. *Jurnal Penelitian Pendidikan IPA*, 8(2), 932–938.
- Hunaepi, H., & Suharta, I. G. P. (2024). Transforming education in Indonesia: The impact and challenges of the Merdeka Belajar curriculum. *Path of Science*, 10(6), 5026–5039.
- Jee, B. D., Matlen, B. J., Greenlaw, M., Simms, N., & Gentner, D. (2022). Spatial supports for comparison in educational science images. *Instructional Science*, 50(6), 807–827.
- Jin, K., Siu, W., & Huang, X. (2022). Exploring the impact of random guessing in distractor analysis. *Journal of Educational Measurement*, 59(1), 43–61.
- Ketterlin-Geller, L. R., McMurrer, J., & Barton, T. (2024). Designing classroom assessments with the end user in mind. *International Journal of Testing*, 24(2), 103–129.
- Khasawneh, Y. J. A., & Khasawneh, M. A. S. (2023). The effectiveness of using infographics on the academic achievement of elementary students. *Migration Letters*, 20(S5), 1258–1271.
- Komalasari, I. S., Rahmi, Y. L., Lufri, & Darussyamsu, R. (2023). Pentingnya mengembangkan instrumen kemampuan berpikir tingkat tinggi tentang materi bioproses sel untuk peserta didik SMA/MA. *Ruang-Ruang Kelas: Jurnal Pendidikan Biologi*, 3(1), 1–6.

- Li, S. (2023). Working memory and second language writing: A systematic review. *Studies in Second Language Acquisition*, 45(3), 647–679.
- Li, T., Zhan, Z., Lin, Z., & Zhang, X. (2026). Assessment of higher-order thinking in education: A systematic review of empirical studies. *Studies in Educational Evaluation*, 90, 101655.
- Lian, L. H. (2023). Assessment of higher-order thinking skills: Is it simply determined by verbs? *International Journal of Academic Research in Progressive Education and Development*, 12(2), 1338–1348.
- Mahmud, M. M., & Wong, S. F. (2022). Digital age: The importance of 21st century skills among the undergraduates. *Frontiers in Education*, 7, 950553.
- Molnár, Z., Kiss, P. S., Gálfi, M., & Radács, M. (2026). The importance of environmental attitudes in dealing with the environmental crisis and the relationship between education systems: A systematic review. *Discover Education*, 5(1), 113.
- Otaya, L. G., Yahiji, K., Rasimin, R., Rahmawati, R., Imtihan, N., & Muhamad, N. B. (2025). Assessment of teachers' ability to develop HOTS-based test items: A GRM analysis. *Jurnal Ilmiah Peuradeun*, 13(3), 1995–2018.
- Purwanto, A., Rahmawati, Y., Rahmayanti, N., Mardiah, A., & Amalia, R. (2022). Socio-critical and problem-oriented approach in environmental issues for students' critical thinking skills development in chemistry learning. *Journal of Technology and Science Education*, 12(1), 50–67.
- Rapti, S., Sapounidis, T., Tselegkaridis, S., & Stamovlasis, D. (2026). Exploring communication, collaboration, critical thinking, and creativity in educational robotics curriculum. *Education and Information Technologies*, 31(9), 2539–2565.
- Risni, T. W., & Khotimah, A. K. (2025). Challenges of fiqh teachers in developing HOTS-based assessments: A case study in Indonesian Islamic senior high school. *Al-Adabiyah: Jurnal Pendidikan Agama Islam*, 6(4), 415–428.
- Ristanto, R. H., Suryanda, A., & Indraswari, L. A. (2023). The development of ecosystem misconception diagnostic test. *International Journal of Evaluation and Research in Education*, 12(4), 2246–2255.
- Riza Andriani, Safriana, Winata, I. W., & Ameliah, I. (2024). Pendampingan penyusunan soal berbasis HOTS untuk guru IPA SMP di Kota Lhokseumawe: Upaya peningkatan kompetensi dan inovasi pembelajaran. *PUSAKA: Jurnal Pengabdian Masyarakat*, 1(2), 7–17.
- Saepuzaman, D., Retnawati, H., Istiyono, E., & Haryanto, H. (2022). Developing HOTS instruments: Is it difficult for physics teachers? *Jurnal Pendidikan MIPA*, 23(4), 1802–1826.
- Samaduri, A. (2022). Analisis pemahaman konsep siswa yang diukur menggunakan tes pilihan ganda beralasan pada mata pelajaran biologi. *Jurnal Pendidikan Glasser*, 6(1), 109–118.
- Sibiç, O., Akçay, B., & Arık, M. (2022). Review of two-tier tests in the studies: Creating a new pathway for development of two-tier tests. *International Journal of Contemporary Educational Research*, 7(2), 81–98.
- Sri Hanipah. (2023). Analisis Kurikulum Merdeka Belajar dalam memfasilitasi pembelajaran abad ke-21 pada siswa menengah atas. *Jurnal Bintang Pendidikan Indonesia*, 1(2), 264–275.
- Tahmid, T., Nurkhoiri, N., & Syaiful, M. (2024). Implementasi perangkat belajar sains berbasis ESD pada materi pencemaran lingkungan untuk peningkatan literasi dan keterampilan berpikir kritis siswa. *Diklabio: Jurnal Pendidikan dan Pembelajaran Biologi*, 8(2), 235–245.

- Thaariq, Z. Z. A. (2026). Bridging theory and context: An integrative review for developing an adaptive curriculum framework in Indonesia. *Review of Education*, 14(2), e70163.
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159–169.
- Uralovich, K. S., Toshmamatovich, T. U., Kubayevich, K. F., Sapaev, I. B., Saylaubaevna, S. S., Beknazarova, Z. F., & Khurramov, A. (2023). A primary factor in sustainable development and environmental sustainability is environmental education. *Caspian Journal of Environmental Sciences*, 21(4), 965–975.
- Zhou, Y., Gan, L., Chen, J., Wijaya, T. T., & Li, Y. (2023). Development and validation of a higher-order thinking skills assessment scale for pre-service teachers. *Thinking Skills and Creativity*, 48, 101272.