International Journal of Education and Practice
2026 Vol. 14, No. 2, pp. 338-359
ISSN(e): 2310-3868
ISSN(p): 2311-6897
DOI: 10.18488/61.v14i2.4898

HOTS-based science questions as assessment instrument for elementary school students: Examining validity and reliability

Endang Susantini[1+]
Julianto[2]
Maryam Isnaini Damayanti[3]
Firas Khaleyla[4]
Mi'rojul Huda[5]

[1,4]*Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya, Indonesia.*
[1]*Email:* endangsusantini@unesa.ac.id
[4]*Email:* firaskhaleyla@unesa.ac.id
[1,2]*EduSTEM Center, Universitas Negeri Surabaya, Indonesia.*
[2,3]*Faculty of Education, Universitas Negeri Surabaya, Indonesia.*
[2]*Email:* julianto@unesa.ac.id
[3]*Email:* maryamdamayanti@unesa.ac.id
[5]*Faculty of Political and Social Sciences, Universitas Negeri Surabaya, Indonesia.*
[5]*Email:* mirojulhuda@unesa.ac.id

(+ Corresponding author)

Article History
Received: 26 September 2025
Revised: 29 January 2026
Accepted: 3 March 2026
Published: 7 April 2026

Keywords
Design and development research
Elementary school students
HOTS-based science questions
Science subjects
Validity and reliability.

ABSTRACT

Assessments based on Higher-order Thinking Skills (HOTS) are critical for elementary students as they are important; however, standardized tools are still lacking. Hence, this study aims to develop valid and reliable HOTS-based science questions as assessment instruments to measure elementary school students' mastery of science subjects and HOTS. This study employed the Design and Development Research (DDR) model to create HOTS-based science questions for grades four, five, and six at the elementary level. The research involved analysis, design, development, and evaluation phases to produce six summative science questions for each grade. Qualitative data were obtained through focus group discussions (FGD) to align with curriculum analysis. Quantitative data were collected via validation sheets and trial tests to assess validity and reliability. The results indicate that there are two science topics with three HOTS-based instructional objectives for each grade. These instructional objectives were developed using three question types: multiple-choice questions (MCQs) with four options, complex MCQs, and essay questions. MCQs were used to evaluate students' analytical skills, complex MCQs to assess evaluative skills, and essay questions to measure creative skills and metacognitive knowledge. The developed HOTS-based science questions were deemed valid (CVR = +0.73, CVI = +1.00) and reliable (ICC = 0.914, Cronbach's $\alpha > 0.80$). This study suggests that the developed HOTS-based science questions can be utilized by elementary science teachers to assess students' mastery of science subjects and HOTS.

Contribution/Originality: This study develops and examines the validity and reliability of HOTS-based science questions based on the Revised Bloom's Taxonomy to be used as a standardized assessment tool in upper levels of elementary school.

1. INTRODUCTION

Higher-order thinking skills (HOTS) become a crucial competence in an era of abundant distorted information and a crisis of criticality caused by the advancement of large language models. In the complex 21st-century landscape, HOTS, including critical thinking, problem-solving, and adaptive application of knowledge, have emerged as necessities for addressing intricate global challenges (Ragab, Kaid, & Sayed, 2024). Organization for Economic Cooperation and Development (OECD) and the World Economic Forum have established HOTS as one of the fundamental competencies critical for developing adaptable individuals in our changing world (OECD, 2019; Schwab & Zahidi, 2020). These skills rise in importance along with the ubiquity of massive language models and internet platforms flooding the web with misinformation, which can lead to dwindling critical thinking (Pratama, Widiati, & Hakim, 2025). Individuals capable in HOTS will flourish; they will no longer be passive recipients of information, but will be able to critically evaluate sources, determine biases, synthesize trustworthy information, and develop novel solutions (Ragab et al., 2024). In addition, HOTS are critical for maintaining intellectual autonomy, so students can distinguish truth from false information and increase critical thinking amidst the era of misinformation caused by a number of unverifiable sources (Setyarini, 2019). By integrating these competencies, particularly in education, younger generations are equipped with the necessary critical skills to differentiate factual and useful knowledge from misinformation.

HOTS has been regarded as a significant part of the Indonesian curriculum, especially in science subjects at the elementary school level. HOTS has been integrated into the education system to build a critical generation from an early age (Pulungan, Toybah, & Suganda, 2021). In this case, HOTS are reflected in three sequential competencies, covering analyzing (C4), evaluating (C5), and creating (C6) (Anderson, 2005; Krathwohl, 2002; Urgo, Arguello, & Capra, 2019). Students gain important cognitive and practical advantages beyond basic knowledge acquisition as they develop these competencies. At the level of C4, students learn to break down scientific information into smaller components, identify patterns in experiments, and recognize potential biases or errors in data, which stimulates deeper curiosity and systematic thinking (Kuncoro, Harini, & Trimono, 2022; Tuela & Palar, 2022). This ability shows them how to approach problems methodically, for example, when they are studying plant growth cycles or comparing weather patterns. At the level of C5, students evaluate the validity of sources, arguments, and scientific claims, enabling them to separate reliable information from misconceptions (Setiawan, Khair, Ratnadi, Hakim, & Istiningsih, 2020). For example, they can critically evaluate the best method for testing a hypothesis. Finally, at the level of C6, students generate original ideas, design innovative experiments, and develop adaptive solutions, such as inventing a simple water filtration system or proposing eco-friendly solutions to reduce waste (Mohamed & Lebar, 2017). These competencies contribute to strengthening scientific literacy, critical reasoning, and creative problem-solving, enabling students to think independently, engage in evidence-based discussions, and apply their knowledge to real-world challenges (Oyewo, Ramaila, & Mavuru, 2022).

To support HOTS-based science learning at elementary school, science teachers should apply both HOTS-based learning models and HOTS-based assessments. By adopting HOTS-oriented teaching strategies (e.g., inquiry-based learning, problem-based learning, or project-based learning), teachers will be able to encourage students to analyze (C4), evaluate (C5), and create (C6) rather than just memorize (C1), understand (C2), and apply (C3) facts (Yusuf, Purwati, Suhirman, & Prayogi, 2025). For example, when students investigate the reason why some objects float while others sink through guided experiments, they practice breaking down concepts (C4), testing hypotheses (C5), and designing solutions (C6). Duran and Dökme (2016) show that inquiry-based methods significantly improve elementary school students' critical thinking skills. Second, HOTS-based assessments ensure that the evaluation matches the intended skills (Ragab et al., 2024). The assessment for the elementary school level may involve multiple-choice questions (MCQs), complex MCQs, and essays to support the construction of the HOTS-based assessment instrument (Maryani, Prasetyo, Wilujeng, Purwanti, & Fitrianawati, 2021). In addition, learning models and assessments must also be aligned; for example, if science teachers use project-based learning but assess only with fact-recall quizzes, students will not develop HOTS. Conversely, HOTS instructional methods without HOTS assessments cause learners to be unprepared. Thus, integrating both ensures cohesive, developmentally appropriate science learning that cultivates versatile thinkers. Yusuf et al. (2025) also noted that different stimuli highly affected the cognitive response to HOTS questions; thus, HOTS assessments must also provide appropriate stimuli for students.

The lack of HOTS-based science questions availability, easily accessible for fourth to sixth-grade students, poses significant challenges to effective science education at the elementary level. Without sufficient access to well-designed higher-order thinking questions, teachers struggle to properly assess and develop students' C4, C5, and C6 skills. In this situation, teachers usually either rely on traditional low-level recall questions (Barral, Ardi-Pastores, & Simmons, 2018) or spend excessive time creating their own HOTS questions, both of which cause their own challenges. Without proper HOTS assessments, students' meaningful understanding of scientific processes or their ability to apply knowledge to different situations will be difficult to assess (Ichsan et al., 2020). Furthermore, the lack of standardized HOTS question banks results in highly varied assessment quality between classrooms, creating inequalities in science education. This gap also potentially disrupts teachers' professional development, as they miss opportunities to learn exemplary formulations of HOTS questions. Ultimately, limited access to grade-appropriate HOTS science questions possibly perpetuates surface-level learning even at a critical stage when students should be developing important scientific reasoning skills to be applied in real-world problem-solving.

In relation to the lack of standardized HOTS-based science questions as assessment instruments for science learning at the elementary school level, this study aims to develop HOTS-based science questions for fourth, fifth, and sixth-grade elementary school students. This study uses curriculum analysis to set the instructional objectives for each science learning level. In addition, science questions are developed based on Anderson (2005) and Krathwohl (2002) HOTS theoretic fundamentals (Anderson, 2005; Krathwohl, 2002; Urgo et al., 2019) to specify the suitability of the HOTS question model, the developed questions are then evaluated for their reliability and validity, ensuring that they meet the standards required for each question, so that these questions can be referred to by elementary school level science teachers. With these efforts, we believe that the study not only points out the importance of HOTS in science learning but also ensures that elementary school students perceive and master HOTS. The research question proposed for this study is: How is the development of the HOTS-based science questions as an assessment instrument for elementary school students?

2. LITERATURE REVIEW

2.1. Higher-Order Thinking Skills (HOTS) in Science Education

Higher-Order Thinking Skills (HOTS) are a level of cognitive functioning that exceeds basic cognitions, such as memorization and understanding. HOTS form the basis of 21st-century skills, allowing individuals to discern complex problems, perform critical analysis, and demonstrate intellectual flexibility in various environments (Hamzah, Hamzah, & Zulkifli, 2022; Kania & Kusumah, 2025; Maxnun, Kristiani, & Sulistyaningrum, 2024). Based on Anderson (2005) and Krathwohl (2002) revision of Bloom's taxonomy, HOTS covers three reciprocal cognitive domains, including analysis (deconstructing information to examine connections and organizational principles), evaluation (formulating sensible assessments based on explicit criteria and contextual relevance), and creation (synthesizing elements to generate novel solutions) (Chandio, Pandhiani, & Iqbal, 2016; Seman, Yusoff, & Embong, 2017). The three domains contribute to a progressive intellectual framework significant for overcoming 21st-century challenges.

Analytical domain (C4) requires students to dissect intricate data into several parts in order to identify underlying correlations, patterns, and structural logic. This is displayed in scientific inquiry when they examine engineering designs to comprehend functional principles or analyze experimental data to determine causative correlation. C4-level skills are also necessary to distinguish relationships and causation when studying issues like climate change (Bejjanki, Pathengay, & Snigdha, 2025). Next, the evaluation (C5) domain demands criterion-based judgment to examine the validity, reliability, and ethical implications of information. Related to this, HOTS supports students in analyzing AI-generated content for biases, evaluating competing scientific theories based on available evidence, or creating a critique of the feasibility of sustainable urban plans against ecological and economic constraints (Li, 2024). Lastly, regarding the creation domain (C6), students synthesized, analyzed components, and evaluated insights into innovative solutions. This transformative process can be seen, for example, when students devise water purification systems using recycled materials, develop algorithms to address community problems, or compose hypotheses to bridge gaps in existing research (Prayogo et al., 2024). As elaborated, the three skills, analysis, evaluation, and creation, are indeed interrelated. Thus, aspects of C4, C5, and C6 skills remain essential in the current learning processes to mirror the scientific method and address ambiguities in our evolving world with evidence-based ingenuity.

In coping with science education, many studies have permeated HOTS in science learning and assessment. For instance, prior studies highlight that mobile learning effectively enhances HOTS and communication skills in science education through collaborative, inquiry-based, and project-based learning approaches. These methods foster critical analysis (C4), evaluation (C5), and creation (C6), aligning with 21st-century demands for problem-solving and adaptability (Anderson, 2005; Krathwohl, 2002; Urgo et al., 2019). However, the studies have not supported HOTS-based pedagogy with valid HOTS-based assessments. Similarly, other studies focus on the use of different learning models rather than providing valid and reliable HOTS-based assessments (Jailani, Sugiman, & Apino, 2017; Sirait, Murniarti, & Sihotang, 2021). For instance, Supriyadi et al. (2022) and Sirait et al. (2021) only focus on the implementation of HOTS-based problem-based learning to enhance students' science literacy skills, while others focus on the use of problem-based learning integrated with local wisdom and HOTS (Annam, Ramdani, Makhrus, & Sarkingobir, 2024; Hikmawati, Suastra, Suma, Sudiatmika, & Rohani, 2021; Papilaya & Tuapattinaya, 2022) HOTS-oriented e-project-based learning (Prihatin & Gunawan, 2024; Rati, Arnyana, Dantes, & Dantes, 2023) and HOTS-based inquiry learning (Mubarok, Suprapto, & Adam, 2019; Qamariyah, Rahayu, Fajaroh, & Alsulami, 2021). Moreover, these studies are mostly undertaken in the higher education setting, not at the elementary level. This phenomenon implies that most studies dealing with HOTS-based science learning focus on learning models used to teach certain science materials or attain specific science skills at higher education levels, and few of them focus on the elementary level.

2.2. HOTS-based Assessment for Science Learning

There are few resources available to create HOTS-based science assessment instruments for elementary school students, and some previous research shows certain limitations in their findings. For example, Sidiq et al. (2021) use a quasi-experimental approach with treatment and control groups to examine the efficacy of introducing elementary children to HOTS-based science questions in order to enhance their critical thinking abilities. The development and validation processes of the HOTS-based evaluations used have not been taken into consideration in this study. These could make one wonder if the learning outcome accurately assesses the HOTS and instructional objectives. Ishartono, Desstya, Prayitno, and Sidiq (2021) emphasized the importance of HOTS-based science questions in their research. However, only 35% of the questions in the HOTS-based assessments created for the study fit important criteria, including various solution paths, interpretative depth, and non-algorithmic problem-solving. A number of other essential components are also lacking. Additionally, several questions do not correspond with the intended learning indicators, which lowers their content validity, and the majority of questions do not include visual examples, which are crucial for elementary kids' comprehension. Anggraeni, Prahani, Suprapto, Shofiyah, and Jatmiko (2023) also created a HOTS-based scientific evaluation for fourth-grade science students at an elementary school. Nonetheless, we point out two primary issues with the HOTS-based evaluations: due to time and comprehension constraints, teachers find it challenging to create high-order cognitive questions, which results in tools that mostly focus on lower-order thinking skills (C1–C3). Furthermore, many assessment instruments are not built with clearer test designs and lack information regarding their alignment with curriculum indicators, which lowers their overall validity and efficacy in gauging actual HOTS competencies. Similar deficiencies are also present in other research, which highlights a lack of discussion regarding the validity and reliability of the assessment tools used to measure science skills and HOTS simultaneously, as well as how to deal with science settings at the elementary school level in grades four through six (Arthadewi, Bayu, & Rati, 2024; Setiawan et al., 2020; Suratmi, Laihat, Asnimar, & Handini, 2020).

By examining the above reviews, it is evident that the literature on the implementation of HOTS in science education at the elementary level emphasizes its fundamental role in fostering analytical, evaluative, and creative thinking. Several studies support the effectiveness of HOTS-oriented pedagogies, such as inquiry-based, project-based, and mobile learning approaches, in enhancing students' critical thinking and problem-solving skills. However, these studies primarily focus on instructional strategies rather than the systematic development and validation of HOTS-based assessment instruments. Additionally, most research is designed for secondary or higher education contexts, with less attention given to elementary-level implementation. This highlights a significant research gap in adapting HOTS-based science assessment instruments to the cognitive developmental stage of younger learners. Existing studies reveal multiple shortcomings in HOTS-based science assessment instruments at the elementary level, including inadequate alignment with curriculum standards, underrepresentation of non-algorithmic and open-ended problem types, and limited focus on validity and reliability testing. Many teachers also report difficulties in preparing such items due to time constraints and lack of expertise, resulting in questions that tend to assess lower-order thinking (C1–C3). These gaps underscore the importance of developing a valid, reliable, and curriculum-aligned HOTS-based science assessment instrument for elementary students to accurately measure both science content mastery and higher-order thinking skills.

3. METHODOLOGY

3.1. Data Collection

This study was a research and development experiment to develop HOTS-based science questions as assessment instruments for elementary school. The Design and Development Research (DDR) model was used in this study, starting from analysis (Phase I), design and development (Phase II), and evaluation (Phase III) (Jaya, 2021). Data were collected during a Focus Group Discussion (FGD) in curriculum analysis, a validation sheet, and a trial test. This study invited three experts on elementary school science education for FGD, and three other experts to validate the developed assessment instrument, all of whom had backgrounds in science education, primary education, and science teaching. This study also involved 120 students from public school Ngenep 2 State Elementary School, East Java, about 40 students from each of the fourth, fifth, and sixth grades. The rationale for considering public schools was due to the present study's expectation to ensure that the developed assessment instruments were compatible for students from all kinds of backgrounds, as found in public schools. The students were further asked to complete the developed HOTS-based science questions as a trial test.

Data collected from FGD were analysed using deductive thematic analysis adapted from Brown (2005) including instructional objectives, HOTS-based learning objectives (level C4-C6 skills), and assessment question types. The quantitative data obtained from the validation sheet were analyzed using the intra-class coefficient (ICC), Lawshe's content validity ratio (CVR), and content validity index (CVI). ICC was used to confirm that the validators shared the agreement level of the scores with the assistance of IBM SPSS 25, revealing inter-rater reliability (IRR) scores. ICC score was categorized based on Koo and Li (2016). The CVR was used to measure whether each of the validation items showed a positive score, whereas the CVI was used to confirm the overall validity judgment (CVI > +0.70, categorized as very high validity) (Almanasreh, Moles, & Chen, 2019). The criteria of validity were calculated based on the following formula.

$$CVR=((ne-N/2))/(N/2)$$

$$CVI=(\sum CVR)/k$$

Note:

CVR: Content Validity Ratio.

ne : Number of raters who considered the item.

N : Total raters.

k : Number of items.

Data from the trial test were then analyzed for their internal consistency reliability (ICR), revealing the Cronbach's α value. The results of the ICR test were considered to indicate very high reliability if the Cronbach's α value met or exceeded 0.80 (Pallant, 2020). To succeed in data collection and analysis, we enclosed a summary of the data, source of data, analysis technique, and expected outcomes used in each phase. Table 1 portrays the data processing for each phase of DDR.

Table 1. Data processing in each phase of DDR.

Phase	Data	Data source	Analysis technique	Expected data outcomes
I: Analysis	Qualitative data of curriculum analysis	FGD notes	Deductive thematic analysis	Determined instructional objectives and HOTS-based learning objectives (cognitive levels) for the science subject at the fourth, fifth, and sixth grades of elementary school.
II: Design and Development	Instructional objectives and HOTS-based learning objectives (cognitive levels)	Results of Phase I	Making sure the relevance between both objectives and the questions	A series of HOTS-based science questions as assessment instruments for science subjects at the fourth, fifth, and sixth grades of elementary school level.
III: Evaluation	- Quantitative data of validity and inter-rater reliability scores - Quantitative data from a trial test	- Validation sheet - Trial test	- ICC test, CVR, and CVI - ICR test	- ICC and Validity judgements - ICR judgment using the Cronbach's α value

3.2. Procedure of DDR

We conducted the three phases of DDR in a sequential order, from the analysis to the evaluation stages. We involved different activities correlated with each phase's objectives (see Table 1). The development process of HOTS-based science assessment instruments for elementary students was carried out through three systematic phases. In Phase I (Analysis), the objective was to identify the instructional objectives and HOTS-based learning objectives (cognitive levels) for science subjects in grades four, five, and six. This phase involved conducting a curriculum analysis through a Focus Group Discussion (FGD) with three science education experts specializing in elementary education. Together with the researchers, they collaboratively determined the relevant objectives that aligned with higher-order thinking skills. In the next Phase II (Design and Development), HOTS-based science questions tailored to each grade level were produced.

The researchers were responsible for developing these assessment items, ensuring that each question reflected analysis, evaluation, or creation skills, consistent with Bloom's HOTS level. Finally, Phase III (Evaluation) aimed to obtain valid and reliable HOTS-based science questions. In this stage, the initial draft of the assessment instruments underwent rigorous validation and reliability testing.

This process included input from a broader range of participants: researchers, three subject matter experts in science education, primary education, and science pedagogy, as well as a trial test involving 120 elementary students as participants. Based on the evaluation, necessary revisions were made to finalize the instrument for classroom implementation. To ensure the integrity and validity of the developed assessment instruments, we implemented a rigorous double-checking process to confirm that the instruments accurately measured the intended constructs.

3.3. Research Ethics

We adhered to ethical research standards by carefully considering and minimizing any potential risks associated with the study. The study was approved by the Ethical Committee of Universitas Negeri Surabaya (No. 036/UN38.III.1/DL.01.02/2025). Since this study partially involved human subjects as primary participants in a trial test, informed consent for general participation was obtained. Moreover, six expert participants were involved in the study as part of the Focus Group Discussion (FGD) and validation process. These experts participated voluntarily, were professionally compensated, and faced no physical or psychological risks from their involvement. All data collected during the study were securely stored in private documentation accessible only to the research team.

4. FINDINGS

4.1. Phase I: Analysis

The undertaken FGD involved three experts and the researchers to discuss the curriculum of the science subject at the elementary school level, especially the fourth, fifth, and sixth grades. The involvement of these three grades was because science is taught at those levels at the elementary level. The FGD lasted two hours, with a synchronous meeting using the Zoom platform due to the impossibility of conducting an offline meeting caused by different geographical locations. Table 2 shows the recaps of the science topic materials, instructional objectives, and HOTS-based learning objectives based on the undertaken curriculum analysis.

Table 2. Results of curriculum analysis.

Grade	Science topic materials	Instructional objectives	HOTS-based learning objectives
4th	1. Objects and their properties	- Connecting the stages of the water cycle with the occurring processes. - Understanding changes in the state of matter in everyday life - Mentioning natural water sources used in everyday life.	LOTS (C1-C3)
		- Explaining the impact of water pollution on the environment - Identifying ways to maintain the availability of clean water. - Explaining how to prevent flooding caused by heavy rainfall	HOTS (C4-C6)
	2. Energy	- Connecting types of energy with examples of their use in everyday life - Comprehending renewable and non-renewable energy sources - Mentioning examples of energy sources used in everyday life.	LOTS (C1-C3)
		- Identifying changes in energy forms in everyday life - Identifying the benefits of renewable energy in everyday life - Explaining how to save energy in everyday life	HOTS (C4-C6)
5th	1. Force	- Connecting types of force with examples of their application in everyday life. - Understanding the effect of force on objects - Mentioning the factors that influence the movement of objects.	LOTS (C1-C3)
		- Identifying the relationship between force and motion in everyday life. - Determining examples of the application of style in everyday life.	HOTS (C4-C6)

Grade	Science topic materials	Instructional objectives	HOTS-based learning objectives
		- Explaining the effect of force on objects and providing examples of its application.	
	2. Ecosystem	- Connecting ecosystem components with their roles in the food chain. - Understanding the impact of environmental damage on the ecosystem - Mentioning examples of interactions between living things in an ecosystem.	LOTS (C1-C3)
		- Identifying the impact of environmental change on ecosystems - Determining how to maintain ecosystem balance - Explaining the impact of human activities on ecosystem balance.	HOTS (C4-C6)
6^{th}	1. Circulatory system	- Matching the circulatory organs with their functions - Understanding the function of arteries and veins - Mentioning the main function of red blood cells	LOTS (C1-C3)
		- Diagnosing the effects of heavy blood loss - Determining the correct statement regarding the circulatory system. - Explaining how to maintain a healthy circulatory system	HOTS (C4-C6)
	2. Force (Exploration)	- Matching the type of style with an example of its application - Understanding the relationship between force and motion - Applying the concept of force in influencing the movement of objects, and being aware of the influence of force on the distance and height of the movement of objects in everyday life.	LOTS (C1-C3)
		- Analyzing factors that influence the speed of movement of an object, in this case, humans, based on training and energy. - Evaluating the influence of weight, surface, and friction on the motion of objects in everyday life. - Producing explanations about the relationship between force, energy, and work in real situations, and explaining how the body uses energy to perform work in two different types of activities.	HOTS (C4-C6)

Table 2 illustrates that the distinction between Lower-Order Thinking Skills (LOTS) and Higher-Order Thinking Skills (HOTS) across various instructional objectives in elementary science subjects is significant. Objectives categorized under LOTS (C1–C3) typically involve memorizing, understanding, and applying skills. For instance, in Grade 4, topics such as "Objects and their properties" and "Energy," LOTS objectives focus on identifying, connecting, and describing observable phenomena like stages of the water cycle or types of energy used daily. This requires basic cognitive engagement. In contrast, HOTS objectives (C4–C6) demand deeper reasoning, critical analysis, and evaluation. For example, within Grade 4, students are expected to explain environmental impacts and propose solutions to water-related issues, which pushes them to analyze cause-and-effect relationships and synthesize knowledge. Similar trends are observed in the instructional objectives conveyed in Grades 5 and 6. Therefore, HOTS-based science questions that would be developed focus on six instructional objectives (from two science materials) for each grade.

Table 2 also portrays a clear evolution of complexity, as evident in the "Force" topic between Grades 5 and 6. In Grade 5, HOTS objectives focused on explaining and identifying applications of force, such as identifying motion based on observable examples. However, in Grade 6, the HOTS objectives were significantly more advanced, where students were asked to analyze, evaluate, and construct explanations about how variables like friction, surface texture, and body energy impacted motion and work. This shift reflected a transition from a basic conceptual stratum to more complex scientific reasoning, integrating physics and physiology. Thus, the HOTS objectives in Grade 6 not only required interpretation of the correlation between concepts but also demanded transferable reasoning to real-world scenarios, supporting the development of critical and scientific thinking.

To support the results of Table 2, HOTS-based instructional objectives were interpreted in terms of assessment question types. Figure 1 presents assessment question types determined from the nexus between instructional objectives and HOTS-based learning objectives (C4-C6).

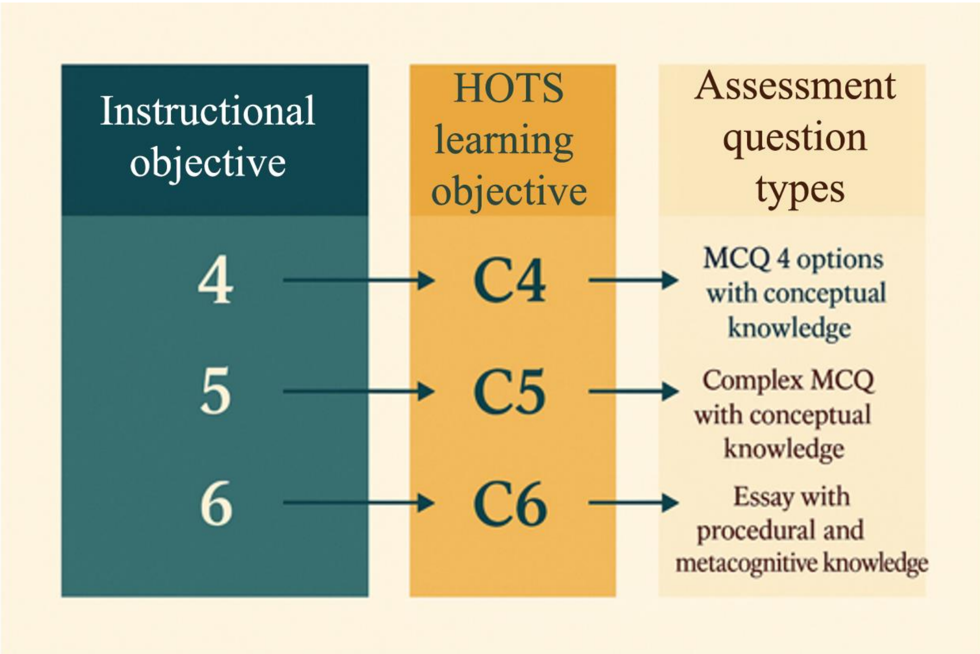

Figure 1. Correlation of instructional objectives, HOTS-based learning objectives, and assessment question types.

Based on Figure 1, the fourth instructional objective at each grade level was designed to foster students' analytical thinking of HOTS. To effectively measure analytical ability (C4), a multiple-choice question (MCQ) with four answer options was designed. This format targeted students' conceptual knowledge, enabling assessment of their ability to differentiate, organize, and relate ideas. The fifth instructional objective focused on evaluation skills (C5), which were assessed using a complex MCQ format. This format required students to process and judge conceptual knowledge critically. Finally, the sixth instructional objective aimed to develop students' creation skills (C6). To capture this advanced cognitive process, questions were designed in an open-ended essay format, emphasizing metacognitive knowledge. This format allowed students to reflect, plan, and generate original responses, making it suitable for evaluating their capacity to create new ideas or solutions. This structured alignment between instructional objectives, cognitive processes, and assessment question types ensured that the assessments meaningfully measured the targeted HOTS components in a valid and pedagogically sound manner.

After obtaining the instructional objectives, HOTS-based learning objectives, and assessment question types, the next phase of designing and developing the HOTS-based science questions is necessary. The following subsection rigorously elaborates on the question model and all related information to ensure compliance with good assessment instruments.

4.2. Phase II: Design and Development

HOTS-based science questions were designed and developed according to the results of Phase I. Before developing the HOTS-based science questions, this study designed the blueprint of the assessment instruments for each grade (Table 3).

Table 3. Blueprint of assessment instruments for Grades 4, 5, and 6.

Grade	Topic materials-Instructional objective-HOTS	Question types	Question number
4	Objects and their properties-Instructional objective 4-C4	MCQ 4 options	1
	Objects and their properties- Instructional objective 5-C5	Complex MCQ	2
	Objects and their properties- Instructional objective 6-C6	Essay	3
	Energy- Instructional objective 4-C4	MCQ 4 options	4
	Energy- Instructional objective 5-C5	Complex MCQ	5
	Energy- Instructional objective 6-C6	Essay	6
5	Force- Instructional objective 4-C4	MCQ 4 options	1
	Force- Instructional objective 5-C5	Complex MCQ	2
	Force- Instructional objective 6-C6	Essay	3
	Ecosystem- Instructional objective 4-C4	MCQ 4 options	4
	Ecosystem- Instructional objective 5-C5	Complex MCQ	5
	Ecosystem- Instructional objective 6-C6	Essay	6
6	Circulatory system- Instructional objective 4-C4	MCQ 4 options	1
	Circulatory system- Instructional objective 5-C5	Complex MCQ	2
	Circulatory system- Instructional objective 6-C6	Essay	3
	Force (Exploration)- Instructional objective 4-C4	MCQ 4 options	4
	Force (Exploration)- Instructional objective 5-C5	Complex MCQ	5
	Force (Exploration)- Instructional Objective 6-C6	Essay	6

Based on the above blueprint, Tables 4, 5, and 6 portray the developed questions for the science subjects at the fourth grade of elementary school level.

Table 4. HOTS-based science questions for Grade 4.

Science topic materials	HOTS-based instructional objectives	HOTS-based science questions
Objects and their properties	Explaining the impact of water pollution on the environment	*Instruction*: What are the impacts of water pollution on living things? *Answers*: A. Water becomes cleaner and healthier. B. Fish and aquatic animals can die. C. Plants grow more fertile. D. No impact on the environment.
	Identifying ways to maintain the availability of clean water	*Instruction*: Choose two ways that can be done to maintain the availability of clean water. *Answer*: o Saving water usage o Throwing garbage into rivers o Planting trees in water catchment areas o Using water excessively
	Explaining how to prevent flooding due to high rainfall	*Instruction*: Explain three ways that can be done to prevent flooding due to high rainfall! *Answer*: (Individual essay portraying the ways of preventing floods)

Science topic materials	HOTS-based instructional objectives	HOTS-based science questions
Energy	Identifying changes in energy forms in everyday life	*Instruction*: What energy changes occur in an electric iron? *Answer*: A. Electrical energy becomes heat energy B. Chemical energy becomes kinetic energy C. Light energy becomes electrical energy D. Electrical energy becomes sound energy
	Identifying the benefits of renewable energy in everyday life	*Instruction*: Choose two benefits of using solar energy: *Answer*: o Reduce fossil fuel use o Increase air pollution o Generate electricity in an environmentally friendly way o Deplete natural resources
	Explaining how to save energy in everyday life	*Instruction*: Explain three ways that can be done to save electrical energy at home!

Regarding Table 4, the six developed science questions were categorized as HOTS-based because they required students to engage in complex cognitive processes aligned with analysis (C4), evaluation (C5), and creation (C6) levels of Bloom's Taxonomy. For example, the questions about the effect of water pollution and changes in energy forms required analytical thinking, as students must examine causative relationships for example, how pollutants damage aquatic ecosystems or how forms of energy are transformed in household tools. Meanwhile, questions asking students to choose appropriate actions (e.g., saving water or using solar energy) support evaluation skills, as students must examine the effectiveness and consequences of various behaviors based on environmental value. Finally, the open-ended questions involved creation skills such as explaining methods to prevent flooding or save electricity at home. Students synthesized knowledge and created new solutions or procedures in response to real-world problems, displaying their ability to make concepts more tangible in different contexts. These tasks required students to answer questions beyond simple recall or comprehension, requiring them to think at higher levels to interpret, justify, or construct responses. Thus, all of the questions not only assessed content understanding but also fostered critical, responsible decision-making aligned with scientific literacy and sustainability goals.

Table 5. HOTS-based science questions for Grade 5.

Science topic materials	HOTS-based instructional objectives	HOTS-based science questions
Force	Identifying the relationship between force and motion in everyday life	*Instruction*: When someone pedals a bicycle, the force that works is…. *Answers*: A. Magnetic force and friction force B. Muscle force and friction force C. Gravitational force and electrical force D. Spring force and friction force
	Determining examples of the application of style in everyday life	*Instruction*: Choose two examples of the application of style in everyday life. *Answer*: o Kicking a ball so that it moves forward o A car moves without being driven by an engine o A magnet attracts metal objects

Science topic materials	HOTS-based instructional objectives	HOTS-based science questions
		○ A rock remains on the ground without any force being applied
	Explaining the effect of force on objects and providing examples of its application.	*Instruction*: Explain how force can change the shape, direction, and speed of an object! Give one example of each. *Answer*: (Individual essay)
Ecosystem	Identifying the impact of environmental change on ecosystems	*Instruction*: What happens if the snake population in an ecosystem decreases drastically? *Answer*: A. Rat population increases B. Rat population decreases C. Eagle population increases D. No change in the ecosystem
	Determining how to maintain ecosystem balance	*Instruction*: Choose two ways that can be done to maintain ecosystem balance. *Answer*: ○ Planting trees to protect animal habitats ○ Dumping industrial waste into rivers ○ Reducing the use of hazardous chemicals ○ Fishing with explosives
	Explaining the impact of human activities on ecosystem balance	*Instruction*: Explain three human activities that can damage the balance of the ecosystem and their impacts. *Answer*: (Individual essay)

According to Table 5, the six developed questions on the topics of Force and Ecosystem in Grade 5 were classified as HOTS-based because they required students to demonstrate analysis, evaluation, and creation skills, as outlined in Bloom's HOTS taxonomy.

For instance, the question about pedaling a bicycle and identifying the working forces demanded analytical thinking (C4), as students must distinguish between different force types and their real-life applications. Similarly, recognizing the consequences of a decreasing snake population required students to analyze ecological relationships and predict chain reactions in food webs.

Questions that asked students to select appropriate actions, such as maintaining ecosystem balance or identifying examples of force, required evaluative thinking (C5), as students must assess which behaviors aligned with scientific principles and environmental sustainability. Finally, the open-ended essay questions, such as explaining how force affects objects or how human activities disrupt ecosystems, required creating skills (C6). These tasks challenged students to generate explanations, propose examples, and construct cause-and-effect reasoning based on conceptual understanding.

Thus, each question was designed to promote deeper cognitive engagement and support critical scientific thinking, essential for developing 21st-century skills in elementary school students.

Table 6. HOTS-based science questions for Grade 6.

Science topic materials	HOTS-based instructional objectives	HOTS-based science questions
Circulatory system	Diagnosing the effects of heavy blood loss	*Instruction*: What happens if someone loses a lot of blood? *Answers*: A. No effect on the body B. The body becomes stronger C. Lack of oxygen and nutrients D. Blood will be replaced automatically
	Determining the correct statement regarding the circulatory system	*Instruction*: Choose the correct statement about the human circulatory system. (You can choose more than one option.) *Answer*: ○ The heart has four main chambers ○ Arteries carry oxygen-rich blood ○ Veins carry oxygen-rich blood ○ Red blood cells function to transport oxygen
	Explaining how to maintain a healthy circulatory system	*Instruction*: Raka is a teenager who often eats fast food, rarely exercises, and often stays up late. Lately, he has been feeling tired easily, and his heartbeat is irregular. His science teacher asked Raka to change his lifestyle so that his circulatory system remains healthy. Based on the case, create a healthy lifestyle plan for Raka that can help maintain the health of his circulatory system. Explain the reasons for each action you suggest! *Answer*: (Individual essay)
Force (Exploration)	Analyzing factors that influence the speed of movement of an object, specifically humans, based on training and energy.	*Instruction*: In the park, there are two children running at different speeds. Ani runs faster because she has been practicing regularly, while Budi runs slower and seems to get tired quickly. After a few meters, Ani arrives at the finish line, while Budi is still far from the finish line. What affects the running speed of Ani and Budi based on this case study?? *Answer*: A. Ani's body mass is lighter, so she can run faster. B. Ani has more energy because she exercises, so she gets there faster. C. Budi wears heavier shoes, so he runs slower. D. Budi runs slower because his body is bigger and stronger.
	Evaluating the influence of weight, surface, and friction on the motion of objects in everyday life.	*Instruction*: Two friends, Rina and Dito, are trying to move two different objects. Rina is trying to push a large wooden table, while Dito is pushing a small plastic table. Rina finds it difficult to move the wooden table, even though she pushes it hard. Meanwhile, Dito can easily move the plastic table, even though he pushes it with the same force. Which factors influence Rina's difficulty in moving the wooden table and Dito's ease in moving the plastic table? Choose the two most appropriate answers! *Answer*: A. The wooden table is heavier than the plastic table, so it requires a greater pushing force to move it. B. The frictional force between the wooden table and the floor is smaller than the frictional force between the plastic table and the floor, so the plastic table moves more easily. C. The wooden table has a rougher surface, so the frictional force between the wooden table and the floor is greater than the plastic table.

Science topic materials	HOTS-based instructional objectives	HOTS-based science questions
		D. The plastic table is lighter, so the pushing force required is smaller to move it.
	Producing explanations about the relationship between force, energy, and work in real situations, and explaining how the body uses energy to do work in two different types of activities.	*Instruction*: In a school, two types of games involve movement: tug-of-war and basketball. In tug-of-war, two teams must pull a rope as hard as they can to make their opponents fall to their side. In basketball, players must throw a ball into a basket while moving. Budi, a student who participates in both games, realizes that in tug-of-war, he gets tired more quickly because he has to work harder to pull the rope. However, when playing basketball, even though he runs, he feels less tired and can run longer. Explain what happens in these two games in relation to the concepts of force, energy, and work. Why does Budi get tired more quickly when playing tug-of-war than when playing basketball? Explain how the concepts of force, energy, and work play a role in these two games, and how Budi's body uses different forces and energy in each game. *Answer*: (Individual essay)

Regarding Table 6, the questions developed on the Circulatory System and Force (Exploration) topics for Grade 6 were categorized as HOTS-based because they stimulated students' higher-level thinking through analysis, evaluation, and creation skills.

For example, the questions requiring students to diagnose the effects of heavy blood loss or analyze factors affecting running speed reflected analytical thinking (C4), as students must interpret physiological or physical relationships and identify cause-and-effect patterns from real-life scenarios. Meanwhile, items that asked students to evaluate statements, such as judging the correctness of circulatory system facts or identifying the most influential factors in object movement, required evaluative thinking (C5). Students must examine evidence, compare alternatives, and determine which responses align with scientific underpinnings.

Most notably, the open-ended questions activated creative thinking (C6), such as creating a lifestyle plan for Raka or explaining differences in energy use between tug-of-war and basketball. These questions required students to synthesize knowledge, apply abstract scientific principles to practical contexts, and generate their own solutions. The scenarios encouraged reflection on personal or observable experiences, enhancing the authenticity of learning. Overall, these questions were purposefully designed to move beyond rote memorization and challenge students to reason scientifically, make judgments, and generate explanations, in line with the goals of fostering higher-order thinking and real-world problem-solving.

After the blueprint was developed into question forms, the developed HOTS-based science questions were required to be tested for their validity and reliability. The following subsection reports the results of IRR, CVR, CVI, and ICR measures.

4.3. Phase III: Evaluation

In this subsection, this study reports two different evaluation stages: the experts' validation results and a test trial result with 120 elementary school students. Moving into the experts' validation results, this study portrayed the degree of score agreement among the three validators using the intra-class coefficient (ICC) test, revealing the inter-rater reliability (IRR) scores. Table 7 shows the results of the ICC test.

Table 7. Results of the ICC test.

Measurement type	ICC	*F*	*p*
Single measures	0.372	11.640	0.000
Average measures	0.914	11.640	0.000

Table 7 indicates the consistency or agreement among raters in evaluating a set of items. The single measures ICC value of 0.372 suggests a fair level of agreement when considering individual raters' scores independently. However, this value was relatively low, meaning there was considerable variation in scoring across individual raters. In contrast, the average ICC value was 0.914, reflecting excellent reliability when the scores from multiple raters were averaged. This indicates that, as a group, the raters demonstrated a high level of consistency. The *F*-value of 11.640 with a *p*-value of 0.000 indicates that the results were statistically significant, confirming that the agreement among raters was not due to chance. Overall, while individual ratings varied, combining ratings significantly improved the reliability of the assessment, supporting the validity of using multiple raters to evaluate HOTS-based science questions.

After confirming that the validators' scores were reliable, the validation test was conducted for the content aspect of the developed HOTS-based science questions. In this case, CVR and CVI were examined to assess the degree of validity (Table 8).

Table 8. Results of the CVR and CVI tests.

Aspects	Validators			CVR
	1	2	3	
Questions are in accordance with the objectives and assessment indicators.	4	3.8	4	+0.33
The questions presented are in accordance with the dimensions of knowledge and the dimensions of cognitive processes.	4	3.6	4	+0.33
The answer choices are homogeneous and logical.	4	4	4	+1.00
The reason for the answer key is logical and in accordance with conceptual truth.	4	4	4	+1.00
The language used is clear and appropriate for elementary school students, so that it is easy for students to understand.	4	4	4	+1.00
CVI				+0.73

According to Table 8, the results of the CVR and CVI show that the developed HOTS-based science questions meet acceptable content validity standards. According to the table, three validators assessed five aspects of the questions. Two aspects yielded a CVR of +0.33, indicating moderate agreement among experts. While not perfect, these values still contributed positively to the overall validity. The other three aspects scored the maximum CVR of +1.00, reflecting full agreement among validators. When averaged, the CVI = +0.73, which exceeded the acceptable threshold of 0.70. This suggests that, overall, the developed HOTS-based science questions are valid and appropriate for assessing students' science knowledge and higher-order thinking skills at the elementary level, with strong agreement particularly in terms of clarity, logical consistency, and conceptual accuracy.

Afterward, the developed HOTS-based science questions were trialed with 40 students from each grade in both public and private schools. The scores obtained from the students were used to calculate the Cronbach's α value through the administration of the ICR test. Table 9 portrays the obtained Cronbach's α value for each grade.

Table 9. Results of the ICR test.

Aspects	*N*	Cronbach's α	Category
HOTS-based science questions for Grade 4	40	0.825	Very high reliability
HOTS-based science questions for Grade 5	40	0.832	Very high reliability
HOTS-based science questions for Grade 6	40	0.855	Very high reliability

The results of the reliability analysis using Cronbach's α show that the HOTS-based science questions developed for elementary students possess very high internal consistency across all grade levels. For Grade 4, the instrument yielded a Cronbach's α of 0.825 based on the responses of 40 students, indicating that the items consistently measure higher-order thinking skills for science topic materials. Similarly, the questions for Grade 5 showed a reliability coefficient of 0.832, also categorized as very high, suggesting that the items are well-aligned and reliably assess the intended constructs for HOTS and science topics in Grade 5. At last, for Grade 6, the Cronbach's α value was slightly higher, namely 0.855, within the very high category. Overall, the high reliability coefficients across all three grade levels demonstrate that the assessment instruments are consistent and appropriate for use in educational practice. Therefore, based on the three phases undertaken to develop HOTS-based science questions for the elementary level, the assessment instruments were said to be valid and reliable.

5. DISCUSSION

This study highlights the necessity for well-structured and sequential instructional objectives appropriate to the scaffolding concept, which moves from lower-order thinking skills (LOTS) to higher-order thinking skills (HOTS). The curriculum analysis conducted showed that each IO in science learning is organized from the stages of memorizing (C1), understanding (C2), applying (C3), analyzing (C4), evaluating (C5), and creating (C6). This sequence aligns with Bloom's cognitive taxonomy, which has been recognized as a basis for improving an individual's cognitive skills. The nature of scaffolding in HOTS-based science learning is important for the gradual evolution of students' thinking skills (Alanazi, Osman, & Halim, 2024). If students are immediately given tasks or questions that require HOTS without a strong LOTS basis, they are at risk of experiencing confusion or learning failure (Alrawili, Osman, & Almuntasheri, 2020). On the contrary, by scaffolding students' cognitive development from basic to complex stages, they not only gain knowledge but also develop critical, analytical, and reflective thinking skills simultaneously (Jayarajah, 2017). This can ultimately increase the effectiveness of the learning process because students can link their prior knowledge to the latest one, solve problems creatively, and make decisions logically (Ansori, Arianto, & Khotimah, 2024). Thus, IO should be structured following the flow from LOTS-based to HOTS-based learning goals as an important foundation in forming complete and sustainable cognitive skills.

In addition, this study found that there are three knowledge types used in the IOs of the science subject at the elementary level, namely conceptual, procedural, and metacognitive knowledge. This study accommodates the three knowledge areas in the developed HOTS-based science questions fairly to the subskills of HOTS: conceptual knowledge with C4 and C5 skills, and procedural and metacognitive knowledge with C6 skills. Conceptual knowledge provides a basis for information and data (Pokojski, Oleksiński, & Pruszyński, 2019; Rittle-Johnson & Siegler, 2022); procedural knowledge allows students to generate procedures of scientific activities (Cooper, 2019); while metacognitive knowledge helps students realize their perceived thinking and reflect on their learning process (Rivas, Saiz, & Ossa, 2022). The use of these three types of knowledge at a high cognitive level is very crucial because HOTS requires complex thinking skills, such as analyzing arguments, evaluating evidence, and producing innovative solutions (Hamzah et al., 2022). The chronological use of knowledge from conceptual to metacognitive aligns with the scaffolding approach, which provides gradual support to students according to their level of cognitive development. Students first acquire conceptual and procedural knowledge before they can think reflectively and independently. Related to this, systematic knowledge is integrated into the instructional objectives as a strategic step to improve HOTS effectively. This ensures that students not only master the knowledge but are also able to apply it in real-life scenarios, aligned with the core of meaningful and transformative learning.

During the development of HOTS-based science questions, operational instructional verbs such as explaining, identifying, diagnosing, and analyzing were used to measure C4 level thinking skills (analysis) in Bloom's Taxonomy. Previous studies highlight that analyzing skills involve the ability to break down information into parts, identify relationships between parts, and explain the logical structure of a concept (Rahman, 2019; Rasheva-Yordanova, Iliev,

& Nikolova, 2018). Thus, the four verbs are considered the most representative in stimulating the analytical thinking process. To measure C5 skills (evaluation), operational instructional verbs such as identifying, determining, and evaluating were used to formulate questions. Based on previous research, level C5 skills are related to the ability to assess arguments, ideas, or procedures based on certain criteria, so these verbs are used to trigger critical reflection and in-depth decision-making processes (Rustam & Priyanto, 2022). Finally, for C6 level skills (creating), this study used explaining and producing as the operational instructional verbs. Previous studies have shown that this level reflects the ability to formulate new ideas, design solutions, or create products based on performed analysis and evaluation (Ichsan & Rahmayanti, 2020). The word "producing" directly encourages students to create something original, while "explaining" supports the creative process by demanding coherent and logical reasoning as the basis for creation. Thus, the selection of these instructional verbs aligns with research-based theory and practice to measure HOTS appropriately.

This study also found that the use of different question formats is appropriate to the cognitive level measured, including MCQ with four options for the C4 level, Complex MCQ for the C5 level, and an open-ended essay for the C6 level. This consideration of question format is based on previous studies, which explain that multiple-choice questions are usually effective for measuring analytical skills because they require students to distinguish, organize, and connect information (Maryani et al., 2021).

Meanwhile, complex MCQs carrying complex stimuli, such as contextual cases, are used to measure evaluation skills (C5), because this question type requires logical reasoning and decision-making based on previously determined criteria (Maryani et al., 2021). For the C6 level, the essay question type was chosen because it requires students to compose, develop, and express new ideas in an original way, which is the core of creative skills (Maryani et al., 2021). The use of MCQ with four options, complex MCQ, and essays is emphasized by many studies as relevant in measuring HOTS in students (Edussuriya, Waduge, Lamawansa, & Samaranayake, 2018; Javaeed, 2018; Suswanti, Solikhin, & Ginting, 2024).

Finally, this study uses the Content Validity Ratio (CVR) and Content Validity Index (CVI) tests to ensure the validity of the developed HOTS-based science questions as assessment instruments, the Intraclass Correlation Coefficient (ICC) to assess inter-rater reliability, and the Internal Consistency Reliability (ICR) to measure the overall reliability of the instrument.

A number of studies developing assessment instruments also adopt these measurements because they can provide more robust information about content relevance and consistency (Rosli, Mokhsein, & Suppian, 2023; Wahyuni, Sarwanto, & Atmojo, 2024). ICC is important to ensure agreement between raters, especially in subjective assessments (Hoyt, 2018), while ICR is used to ensure that all instrument items work consistently in measuring the same competency (Pratama, Imaniar, Lestari, Anam, & Rochmawati, 2024). The novelty of this study lies especially in the combination of measurements used to validate the assessment products developed.

In conclusion, while the assessment instruments established in this work have been shown to be valid and trustworthy, several limitations must be acknowledged. In practice, teachers may need to adjust questions based on the demands of the classroom environment, student characteristics, or special learning objectives. The HOTS question model developed has the potential to be utilized as an international reference, particularly for elementary school pupils, although its use will vary depending on the curriculum background of each country. Furthermore, this study is restricted to three question types: essay, difficult multiple-choice, and multiple-choice with four possibilities. Therefore, further evaluation is needed if the instrument is to be applied in other question forms such as short answer, matching, or performance-based assessment. At last, this study only focuses on the development of assessment instruments and does not cover HOTS-based teaching strategies. For this reason, forthcoming researchers are advised to design a standardized HOTS-based science learning model so that it supports the assessment instruments that have been developed.

6. CONCLUSION

This study develops HOTS-based science questions as assessment instruments using the Design and Development Research (DDR) model, consisting of three phases: analysis, design and development, and evaluation. In the analysis phase, this study examines the instructional objectives in science learning at the elementary level. This study found that the instructional objectives have been arranged starting from the lower-order thinking skills (LOTS) to the higher-order thinking skills (HOTS). This shows that the undertaken science learning has adopted a scaffolding approach, where science education guides students from familiarizing themselves with science content to creating ideas or solutions to science problems. The HOTS-based instructional objectives cover analytical, evaluative, and creative thinking processes, demanding students to conceive conceptual, procedural, and metacognitive knowledge, respectively. In the design and development phase, this study uses three assessment question types, namely multiple-choice questions with four options, complex multiple-choice questions, and essays, to accommodate the learning objectives of science topic materials and HOTS. In the evaluation phase, this study conducted rigorous validity and reliability tests to ensure that the developed HOTS-based science questions were consistent with the curriculum contents and C4, C5, and C6 skills. This study is limited by the sample size as participants came only from one public school. In the future, we aim to expand the tested participants to be able to capture more representative data from a wider area.

Funding: This study was funded by Indonesian National Research and Innovation Agency (BRIN) with Indonesia Endowment Fund for Education Agency (LPDP) (Grant number: 37/IV/KS/02/2025) and Universitas Negeri Surabaya (Grant number: 27760/UN38.III.1/KS.03.02/2025).
Institutional Review Board Statement: This study was approved by the Institutional Review Board of LPPM Universitas Negeri Surabaya under protocol number (036/UN38.III.1/DL.01.02/2025). Informed verbal consent was obtained from all participants and all data were anonymized to protect confidentiality.
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

REFERENCES

Alanazi, A. A., Osman, K., & Halim, L. (2024). Effect of scaffolding strategies and guided discovery on higher-order thinking skills in physics education. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*(9), em2496. https://doi.org/10.29333/ejmste/14980

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, *15*(2), 214-221. https://doi.org/10.1016/j.sapharm.2018.03.066

Alrawili, K. S., Osman, K., & Almuntasheri, S. (2020). Effect of scaffolding strategies on higher-order thinking skills in science classroom. *Journal of Baltic Science Education*, *19*(5), 718-729. https://doi.org/10.33225/jbse/20.19.718

Anderson, L. W. (2005). Objectives, evaluation, and the improvement of education. *Studies in Educational Evaluation*, *31*(2-3), 102-113. https://doi.org/10.1016/j.stueduc.2005.05.004

Anggraeni, D. M., Prahani, B. K., Suprapto, N., Shofiyah, N., & Jatmiko, B. (2023). Systematic review of problem based learning research in fostering critical thinking skills. *Thinking Skills and Creativity*, *49*, 101334. https://doi.org/10.1016/j.tsc.2023.101334

Annam, S., Ramdani, A., Makhrus, M., & Sarkingobir, Y. (2024). Development and validation of e-modules integrated with PBL and local wisdom to enhance critical thinking and problem-solving skills. *Pedagogy Review*, *3*(2), 104-115. https://doi.org/10.61436/pedagogy/v3i2.pp104-115

Ansori, I., Arianto, F., & Khotimah, K. (2024). The role of mobile learning in facilitating science learning to enhance higher order thinking skills (HOTS): A systematic review. *Indonesian Journal of Instructional Media and Model*, *6*(2), 79-88. https://doi.org/10.32585/ijimm.v6i2.5580

Arthadewi, R. I., Bayu, I. G. W., & Rati, N. W. (2024). HOTS-based science learning outcomes assessment instrument for measuring elementary school learning outcomes. *Mimbar Ilmu*, *29*(1), 156-162. https://doi.org/10.23887/mi.v29i1.36521

Barral, A. M., Ardi-Pastores, V. C., & Simmons, R. E. (2018). Student learning in an accelerated introductory biology course is significantly enhanced by a flipped-learning environment. *CBE—Life Sciences Education*, *17*(3), ar38. https://doi.org/10.1187/cbe.17-07-0129

Bejjanki, K. M., Pathengay, A., & Snigdha, S. (2025). Commentary: Bloom's taxonomy as an assessment tool in ophthalmology fellowship. *Indian Journal of Ophthalmology*, *73*(Suppl 1), S175. https://doi.org/10.4103/IJO.IJO_777_24

Brown, S. (2005). Assessment for learning. *Learning and Teaching in Higher Education*(1), 81–89.

Chandio, M. T., Pandhiani, S. M., & Iqbal, R. (2016). Bloom's Taxonomy: Improving assessment and teaching-learning process. *Journal of Education and Educational Development*, *3*(2), 203-221.

Cooper, T. (2019). Calling out 'alternative facts': Curriculum to develop students' capacity to engage critically with contradictory sources. *Teaching in Higher Education*, *24*(3), 444-459. https://doi.org/10.1080/13562517.2019.1566220

Duran, M., & Dökme, İ. (2016). The effect of the inquiry-based learning approach on student's critical-thinking skills. *Eurasia Journal of Mathematics Science and Technology Education*, *12*(12), 2887-2908. http://doi.org/10.12973/eurasia.2016.02311a

Edussuriya, D. H., Waduge, R. N., Lamawansa, M. D., & Samaranayake, A. N. (2018). Evaluation of the cognitive level of essay questions of an undergraduate medical program in Sri Lanka, using bloom's taxonomy. *Sri Lanka Journal of Medicine*, *27*(1), 3-10. https://doi.org/10.4038/sljm.v27i1.72

Hamzah, H., Hamzah, M. I., & Zulkifli, H. (2022). Systematic literature review on the elements of metacognition-based higher order thinking skills (HOTS) teaching and learning modules. *Sustainability*, *14*(2), 813. https://doi.org/10.3390/su14020813

Hikmawati, H., Suastra, I. W., Suma, K., Sudiatmika, A. A. I. A. R., & Rohani, R. (2021). The effect of problem-based learning integrated local wisdom on student HOTS and scientific attitude. *Jurnal Penelitian Pendidikan IPA*, *7*(SpecialIssue), 233-239. https://doi.org/10.29303/jppipa.v7iSpecialIssue.1118

Hoyt, W. T. (2018). Interrater reliability and agreement. In M. E. Williams & R. F. B. Turner (Eds.), The reviewer's guide to quantitative methods in the social sciences. In (pp. 132–144). New York: Routledge.

Ichsan, I. Z., Hasanah, R., Ristanto, R. H., Rusdi, R., Cahapay, M. B., Widiyawati, Y., & Rahman, M. M. (2020). Designing an innovative assessment of HOTS in the science learning for the 21st century. *Jurnal Penelitian dan Pembelajaran IPA*, *6*(2), 211-224. https://doi.org/10.30870/jppi.v6i2.4765

Ichsan, I. Z., & Rahmayanti, H. (2020). HOTSEP: Revised Anderson's taxonomy in environmental learning of COVID-19. *European Journal of Educational Research*, *9*(3), 1257-1265. https://doi.org/10.12973/eu-jer.9.3.1257

Ishartono, N., Desstya, A., Prayitno, H. J., & Sidiq, Y. (2021). The quality of HOTS-based science questions developed by Indonesian elementary school teachers. *Journal of Education Technology*, *5*(2), 236-245. https://doi.org/10.23887/jet.v5i2.33813

Jailani, J., Sugiman, S., & Apino, E. (2017). Implementing the problem-based learning in order to improve the students' HOTS and characters. *Jurnal Riset Pendidikan Matematika*, *4*(2), 247-259. https://doi.org/10.21831/jrpm.v4i2.17674

Javaeed, A. (2018). Assessment of higher ordered thinking in medical education: Multiple choice questions and modified essay questions. *MedEdPublish*, *7*, 128. https://doi.org/10.15694/mep.2018.0000128.1

Jaya, I. G. N. (2021). Developing HOTS-based science questions as assessment instruments for elementary school: A design and development research approach. *Journal of Educational Research and Development*, *15*(3), 123-145.

Jayarajah, K. (2017). The development of a year five children's engineering teaching module for HOTS. Doctoral Dissertation, University of Malaya, Malaysia.

Kania, N., & Kusumah, Y. S. (2025). The measurement of higher-order thinking skills: A systematic literature review. *Malaysian Journal of Learning and Instruction*, *22*(1), 97-116.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, *41*(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2

Kuncoro, K. S., Harini, E., & Trimono, D. A. (2022). Bloom's taxonomy analyze category: The analysis of students' analytical skills based on gender. *Unnes Journal of Mathematics Education*, *11*(2), 156-165. https://doi.org/10.15294/ujme.v11i2.58473

Li, T. (2024). Cognitive synergy: Exploring the transformative intersection of human intelligence and artificial intelligence in designing equitable next generation science assessments. Doctoral Dissertation, Michigan State University. Michigan State University Digital Repository.

Maryani, I., Prasetyo, Z. K., Wilujeng, I., Purwanti, S., & Fitrianawati, M. (2021). HOTS multiple choice and essay questions: A validated instrument to measure higher-order thinking skills of prospective teachers. *Journal of Turkish Science Education*, *18*(4), 674-690. https://doi.org/10.36681/tused.2021.97

Maxnun, L. L., Kristiani, K., & Sulistyaningrum, C. D. (2024). Development of HOTS-based cognitive assessment instruments: ADDIE model. *Journal of Education and Learning (EduLearn)*, *18*(2), 489-498. https://doi.org/10.11591/edulearn.v18i2.21079

Mohamed, R., & Lebar, O. (2017). Authentic assessment in assessing higher order thinking skills. *International Journal of Academic Research in Business and Social Sciences*, *7*(2), 466-476.

Mubarok, H., Suprapto, N., & Adam, A. S. (2019). Using inquiry-based laboratory to improve students' higher order thinking skills (HOTs). *Journal of Physics: Conference Series*, *1171*(1), 012040. https://doi.org/10.1088/1742-6596/1171/1/012040

OECD. (2019). An OECD learning framework 2030. In Bast, G., Carayannis, E.G., Campbell, D. F. J. (Eds.), The Future of Education and Labor. Arts, Research, Innovation and Society. In (pp. 23–35). Cham: Springer International Publishing.

Oyewo, O. A., Ramaila, S., & Mavuru, L. (2022). Harnessing project-based learning to enhance STEM students' critical thinking skills using water treatment activity. *Education Sciences*, *12*(11), 780. https://doi.org/10.3390/educsci12110780

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7th ed.). London, UK: Routledge.

Papilaya, P. M., & Tuapattinaya, P. M. J. (2022). Problem-based learning dan creative thinking skills students based on local wisdom in Maluku. *Al-Ishlah: Jurnal Pendidikan*, *14*(1), 429-444. https://doi.org/10.35445/alishlah.v14i1.1406

Pokojski, J., Oleksiński, K., & Pruszyński, J. (2019). Knowledge based processes in the context of conceptual design. *Journal of Industrial Information Integration*, *15*, 219-238. https://doi.org/10.1016/j.jii.2018.07.002

Pratama, R. D., Imaniar, F., Lestari, L. A., Anam, S. U., & Rochmawati, L. (2024). EFL teacher self and collective efficacy scale (ETSCE): Developing efficacy measures in teaching EFL. *Englisia: Journal of Language, Education, and Humanities*, *11*(2), 161-180. https://doi.org/10.22373/ej.v11i2.22232

Pratama, R. D., Widiati, U., & Hakim, L. N. (2025). Effects of human-AI collaborative writing strategy on EFL students' argumentative writing skills. *Computer Assisted Language Learning*, 1-32. https://doi.org/10.1080/09588221.2025.2503900

Prayogo, W., Ihsan, M. F., Awan, F. N., Mawaddah, N., Khairuni, Z. I., Wibowo, H., . . . Lukman, M. F. (2024). A review of sustainable water practices: Teaching high school students to manage and purify water for daily needs. *Indonesian Journal of Social and Environmental Issues*, *5*(3), 287-298. https://doi.org/10.47540/ijsei.v5i3.1724

Prihatin, P., & Gunawan, W. (2024). The influence of E-PjBL and student learning interest on critical thinking skills. *Jurnal Inovasi dan Teknologi Pembelajaran*, *11*(1), 24-32. https://doi.org/10.17977/um031v11i12024p024

Pulungan, M., Toybah, T., & Suganda, V. A. (2021). Development of HOTS-based 2013 curriculum assessment instruments in elementary school. *Journal of Teaching and Learning in Elementary Education*, *4*(1), 51-64.

Qamariyah, S. N., Rahayu, S., Fajaroh, F., & Alsulami, N. M. (2021). The effect of implementation of inquiry-based learning with socio-scientific issues on students' higher-order thinking skills. *Journal of Science Learning*, *4*(3), 210-218. https://doi.org/10.17509/jsl.v4i3.30863

Ragab, A., Kaid, A., & Sayed, A. K. (2024). Enhancing higher order thinking skills (HOTS) in education: Strategies and outcomes. *TOFEDU: The Future of Education Journal*, *3*(5), 1488-1499.

Rahman, M. M. (2019). 21st century skill "problem solving": Defining the concept. *Asian Journal of Interdisciplinary Research*, *2*(1), 64-74. https://doi.org/10.34256/ajir1917

Rasheva-Yordanova, K., Iliev, E., & Nikolova, B. (2018). *Analytical thinking as a key competence for overcoming the data science divide.* Paper presented at the EDULEARN18 Proceedings, IATED.

Rati, N. W., Arnyana, I. B. P., Dantes, G. R., & Dantes, N. (2023). HOTS-oriented e-project-based learning: Improving 4C skills and science learning outcome of elementary school students. *International Journal of Information and Education Technology*, *13*(6), 959-968. https://doi.org/10.18178/ijiet.2023.13.6.1892

Rittle-Johnson, B., & Siegler, R. S. (2022). The relation between conceptual and procedural knowledge in learning mathematics: A review. In J. C. Smith & L. M. Jones (Eds.), The development of mathematical skills. In (pp. 75–110). New York: Academic Press.

Rivas, S. F., Saiz, C., & Ossa, C. (2022). Metacognitive strategies and development of critical thinking in higher education. *Frontiers in Psychology*, *13*, 913219. https://doi.org/10.3389/fpsyg.2022.913219

Rosli, R., Mokhsein, S. E., & Suppian, Z. (2023). Content validity of teachers' efficacy instruments toward implementing of classroom assessment in primary schools using content validity ratio (CVR). *Journal of Contemporary Social Science and Education Studies*, *3*(2), 22-31. https://doi.org/10.5281/zenodo.10154858

Rustam, R., & Priyanto, P. (2022). Critical thinking assessment in the teaching of writing Indonesian scientific texts in high school. *Jurnal Penelitian dan Evaluasi Pendidikan*, *26*(1), 12-25. https://doi.org/10.21831/pep.v26i1.36241

Schwab, K., & Zahidi, S. (2020). *The global competitiveness report 2020: How countries are performing on the road to recovery*. Geneva, Switzerland: World Economic Forum.

Seman, S. C., Yusoff, W. M. W., & Embong, R. (2017). Teachers challenges in teaching and learning for higher order thinking skills (HOTS) in primary school. *International Journal of Asian Social Science*, *7*(7), 534-545. https://doi.org/10.18488/journal.1.2017.77.534.545

Setiawan, H., Khair, B. N., Ratnadi, R., Hakim, M., & Istiningsih, S. (2020). *Developing HOTS-based assessment instrument for primary schools*. Paper presented at the 1st Annual Conference on Education and Social Sciences (ACCESS 2019), Atlantis Press.

Setyarini, S. (2019). *Promoting learner autonomy in teaching English to young adolescents through higher order thinking skills: An innovation to create 4.0 learning trends*. Paper presented at the 1st International Conference on Education Social Sciences and Humanities (ICESSHum 2019), Atlantis Press.

Sidiq, Y., Ishartono, N., Desstya, A., Prayitno, H. J., Anif, S., & Hidayat, M. L. (2021). Improving elementary school students' critical thinking skill in science through HOTS-based science questions: A quasi-experimental study. *Jurnal Pendidikan IPA Indonesia*, *10*(3), 378-386. https://doi.org/10.15294/jpii.v10i3.30891

Sirait, S., Murniarti, E., & Sihotang, H. (2021). Implementation of HOTS-based learning and problem based learning during the pandemic of COVID-19 in SMA Budi Mulia Jakarta. *Advances in Social Sciences Research Journal*, *8*(2), 296-305. https://doi.org/10.14738/assrj.82.9727

Supriyadi, S., Tias, I. W. U., Astuti, N., Sukamto, I., Destini, F., Khairani, F., & Izzatika, A. (2022). Implementation of HOTS-oriented problem-based learning on science literacy ability. *Jurnal Pendidikan Progresif*, *12*(3), 1492-1499. https://doi.org/10.23960/jpp.v12.i3.202236

Suratmi, S., Laihat, L., Asnimar, A., & Handini, E. O. (2020). *Teachers' understanding of HOTS-based assessment in elementary schools*. Paper presented at the Proceedings of the International Conference on Elementary Education (ICEE).

Suswanti, N., Solikhin, F., & Ginting, S. M. (2024). Development of cognitive assessment instruments on buffer solution material based on HOTS (higher order thinking skills). *Journal of Chemistry Education Research*, *8*(2), 97-103. https://doi.org/10.26740/jcer.v8n2.p97-103

Tuela, A. I., & Palar, Y. N. (2022). Analysis of higher order thinking skills (HOTS) based on bloom taxonomy in comprehensive examination questions. *Jurnal Kependidikan: Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran, dan Pembelajaran*, *8*(4), 957-971. https://doi.org/10.33394/jk.v8i4.5885

Urgo, K., Arguello, J., & Capra, R. (2019). *Anderson and krathwohl's two-dimensional taxonomy applied to task creation and learning assessment*. Paper presented at the Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval (pp. 117-124). https://doi.org/10.1145/3341981.3344226

Wahyuni, L., Sarwanto, S., & Atmojo, I. R. W. (2024). Measurement of science literacy skills of elementary school teacher education students: Development and validity testing of assessment instruments. *International Journal of Current Science Research and Review*, 7(11), 8416–8422. https://doi.org/10.47191/ijcsrr/V7-i11-27

Yusuf, Purwati, N., Suhirman, & Prayogi, S. (2025). The effect of HOTS-based question stimulus types and learning styles on students' cognitive response, thoroughness, and conceptual understanding in biology learning. *International Journal of Education and Practice*, *13*(3), 1143-1166. https://doi.org/10.18488/61.v13i3.4372