ISEJ Indonesian Science Education Journal

ISSN Online: 2716-3350
https://siducat.org/index.php/isej/
Published by Yayasan Darussalam Bengkulu


# Development and Validation of a Critical Thinking Assessment Instruments for Use with Sc-SEBION Learning Media

**Henik Anjayati[1], Yuni Sri Rahayu[2], Yuliani[3]**
Master's Program in Biology Education, Universitas Negeri Surabaya, Surabaya, Indonesia[1,2,3]
E-mail Corresponding: hanjayati.biotechnology@gmail.com

| Received: Februari 17, 2026 | Revised: May 4, 2026 | Accepted: May 25, 2026 |
|---|---|---|

**Abstract**
Critical thinking is an essential competency in twenty-first-century education. However, Indonesian students' critical thinking skills remain relatively low, highlighting the need for innovative learning media that support higher-order thinking. Sc-SEBION (*Syzygium cumini*–Explosion Box Interactive Online Version) was developed as an interactive digital learning medium to facilitate critical thinking through contextual and engaging learning experiences. To evaluate its effectiveness, a valid and reliable assessment instrument is required. This study aimed to determine the validity and reliability of a cognitive assessment instrument developed to measure Grade VII junior high school students' critical thinking skills on environmental change topics. The instrument, consisting of pretest and posttest forms, was developed using the 4D model (Define, Design, Develop, and Disseminate). Both forms comprised 12 essay items representing six critical thinking indicators proposed by Facione: interpretation, analysis, evaluation, inference, explanation, and self-regulation. Specifically, items 1–2 measure interpretation, items 3–4 measure analysis, items 5–6 measure evaluation, items 7–8 measure inference, items 9–10 measure explanation, and items 11–12 measure self-regulation. The pretest and posttest were parallel forms with different items but identical constructs, blueprint, and weighting, indicating equivalence in content and structure. The instrument was then tested for validity and reliability. Validity tests consisted of content validity and empirical validity. Content validity was evaluated by four validators, comprising three university lecturers and one science teacher, while empirical validity and reliability were analyzed using Pearson item–total correlation and Cronbach's Alpha in SPSS version 29. Empirical validity and reliability were tested on 21 students outside the research sample. Content validity scores ranged from 91.7% to 96.5%. Empirical validity coefficients ranged from 0.606 to 0.927 for the pretest and 0.657 to 0.916 for the posttest, exceeding the critical r-table of 0.433. Cronbach's Alpha values were 0.936 and 0.934 for the pretest and posttest, respectively. The findings provide preliminary evidence of the instrument's validity and internal consistency for assessing critical thinking skills in Sc-SEBION-based learning. Given the limited pilot sample, further studies with larger and more diverse samples and additional construct validation analyses are needed to strengthen the psychometric evidence.

**Keywords:** Cognitive Assessment Instrument; Critical Thinking Skills; Validity; Reliability; Sc-SEBION.

## INTRODUCTION

Twenty-first-century education increasingly requires students to navigate complex information, evaluate competing claims, interpret evidence, communicate reasoned judgments, and participate responsibly in scientific and civic decision-making. Within this landscape, critical thinking has become a central educational priority because it enables learners to move beyond factual recall toward reflective judgment, problem analysis, evidence appraisal, inference, explanation, and self-regulation. The prominence of critical thinking is closely associated with the broader framework of twenty-first-century competencies, commonly represented through the 4Cs: critical thinking, creativity, communication, and collaboration (Dwiprabowo et al., 2024). Ennis (2011) conceptualized critical thinking as reasonable and reflective thinking

focused on deciding what to believe or do, whereas Facione (1990) emphasized interpretation, analysis, evaluation, inference, explanation, and self-regulation as core dimensions of critical thinking. More recent scholarship has further demonstrated that critical thinking is essential for students' participation in knowledge-based, digitally mediated, and increasingly uncertain societies (OECD, 2023; Rodríguez-Rojas et al., 2024; van Laar et al., 2020).

The significance of critical thinking is particularly evident in science education, where students are expected not only to understand scientific concepts but also to interpret phenomena, evaluate evidence, recognize causal relationships, and formulate evidence-based solutions to authentic problems. Contemporary science education therefore needs to facilitate intellectual engagement with real-world issues, including climate change, biodiversity decline, environmental degradation, food security, and sustainable development. These issues are inherently complex because they involve scientific evidence, social consequences, ethical considerations, and uncertainty. Students must consequently be prepared to interpret information critically rather than accept simplified explanations without scrutiny. International studies have consistently indicated that inquiry-oriented, project-based, and technology-supported learning environments can support the development of critical thinking when students are required to investigate meaningful problems, evaluate alternative explanations, and justify their conclusions using evidence (Almulla, 2020; Cortázar et al., 2021; Guo et al., 2020).

Despite the acknowledged importance of critical thinking, its development remains a persistent challenge in many educational systems, including Indonesia. The Programme for International Student Assessment has repeatedly highlighted the need to strengthen students' capacity to apply knowledge, reason with evidence, and solve complex problems in unfamiliar contexts (OECD, 2019, 2023). Although PISA does not measure critical thinking as a single isolated construct, its literacy, mathematical, and scientific reasoning tasks require competencies closely connected with critical interpretation, analytical reasoning, and evidence-based judgment. In Indonesia, this challenge is also reflected in studies reporting uneven critical-thinking performance across science-learning contexts, particularly when students are asked to analyze complex phenomena, evaluate arguments, or formulate reasoned explanations rather than reproduce memorized information (Primasari et al., 2020; Ramdani et al., 2021; Tanti et al., 2020). Therefore, improving critical thinking requires not only innovative instructional approaches but also assessment instruments that can validly capture students' higher-order cognitive performance.

The development of critical thinking is increasingly intertwined with digital transformation in education. Digital media can potentially expand access to multimodal information, interactive simulations, visual representations, collaborative spaces, and self-paced exploration. Nevertheless, digitalization alone does not automatically produce deeper learning. Technology becomes pedagogically meaningful only when its features are deliberately aligned with learning objectives, task structures, feedback processes, and assessment practices. Studies have shown that technology-supported learning can strengthen engagement, inquiry, and critical thinking when students are actively required to process information, articulate arguments, compare evidence, and reflect on their learning strategies (Isnaeni et al., 2021; Perry et al., 2022; van Laar et al., 2020). Conversely, digital media may remain superficial when students merely consume content without undertaking analytical, evaluative, or reflective activities. Thus, the quality of digital learning should be assessed not merely through technological attractiveness but through its capacity to support meaningful cognitive engagement.

Interactive and project-oriented digital environments are particularly promising because they position students as active participants in constructing knowledge. Project-based learning, for example, can promote student engagement when learners identify authentic problems, investigate relevant evidence, collaborate with peers, develop possible solutions, and communicate their findings. A review by Guo et al. (2020) indicated that project-based learning may improve cognitive, affective, and collaborative outcomes when supported by clear instructional structures and appropriate assessment. Similarly, Almulla (2020) emphasized that

project-based learning can enhance student engagement because it connects learning tasks with purposeful inquiry and practical outcomes. In online contexts, Cortázar et al. (2021) found that structured project-based learning can foster critical thinking when students receive scaffolding, feedback, and opportunities for collaborative argumentation. These studies imply that digital learning media should be designed not merely as repositories of information but as structured environments that prompt students to interpret, analyze, evaluate, and communicate their understanding.

In the Indonesian context, the need for visually engaging and contextually meaningful digital media has become increasingly evident. A needs analysis by Zulaiha et al. (2025) indicated that junior high school students require more attractive and interactive learning media to support their understanding of environmental issues. Similarly, Indonesian research has shown that digital learning media, interactive e-modules, and inquiry-based applications can facilitate critical-thinking development when they provide students with opportunities to explore scientific phenomena, examine relationships between concepts, and respond to challenging questions (Isnaeni et al., 2021; Sujanem & Suwindra, 2023; Verawati et al., 2021). These findings suggest that digital media should be embedded within a pedagogical design that explicitly targets higher-order thinking rather than simply presenting visual or multimedia content.

One form of digital innovation relevant to this need is Sc-SEBION (*Syzygium cumini*–Explosion Box Interactive Online Version), an interactive digital learning medium inspired by the explosion box concept. Sc-SEBION is designed to offer layered, visually engaging, and exploratory learning experiences through digital content, interactive features, and contextually organized environmental-change materials. Unlike conventional printed learning materials, the digital explosion box format allows students to access multiple representations of a topic through visual stimuli, structured information, interactive elements, and guided activities. Previous research on explosion-box media has suggested that such media can stimulate curiosity, increase learner engagement, and support more concrete understanding when instructional content is carefully organized (Fatmawati et al., 2024; Nasriyah et al., 2021; Safitri, 2024). However, the educational value of this media depends on whether it can encourage students to engage intellectually with the content rather than merely interact with its visual features.

Sc-SEBION is distinctive because it integrates local wisdom related to *juwet* (*Syzygium cumini*) into environmental-change learning. Juwet Lopang is a distinctive plant associated with Sumberejo–Lopang, Kembangbahu, Lamongan, and has been connected to local cultural and socio-economic practices (Ashkar et al., 2022). Nevertheless, awareness of its ecological and cultural significance has gradually declined, while environmental pressures such as land conversion, ecosystem disturbance, and climate-related changes may threaten plant populations and local biodiversity. Integrating *juwet* into science learning is therefore not merely a contextual illustration; it offers students an opportunity to connect abstract environmental concepts with ecological realities that are socially, culturally, and geographically meaningful. Contextualization is important because students are more likely to engage in deeper reasoning when scientific issues are connected to familiar experiences and observable problems within their communities.

The educational integration of local wisdom has gained attention because it can bridge formal scientific knowledge with students' cultural experiences and environmental realities. Local-wisdom-based science education may improve relevance, engagement, and conceptual understanding by positioning learners as members of communities with particular ecological histories and practices. In Indonesia, studies have reported that ethnoscience, Ethno-STEM, and local-wisdom-based learning can contribute to critical thinking, creativity, scientific literacy, and problem-solving skills when cultural resources are connected systematically with scientific concepts and inquiry activities (Martawijaya et al., 2023; Ramdani et al., 2021; Sumarni & Kadarwati, 2020). Furthermore, environmental problem-based learning materials that draw on local contexts can strengthen students' environmental literacy by encouraging them to analyze environmental challenges and consider feasible conservation responses (Suryawati et al., 2020; Zulaiha et al., 2025; Amanda et al., 2025).

However, contextual digital learning media alone are insufficient without valid instruments capable of measuring the intended learning outcomes. The implementation of Sc-SEBION requires an assessment system that can identify whether students demonstrate critical thinking while engaging with environmental-change content, local biodiversity issues, and project-based learning activities. Assessment should therefore move beyond factual recall and examine how students interpret environmental phenomena, analyze causal relationships, evaluate evidence, formulate inferences, explain their reasoning, and regulate their own thinking (Bhakti et al., 2025; Rodríguez-Rojas et al., 2024; Verawati et al., 2020). Such an approach is consistent with Facione's critical-thinking framework and with contemporary assessment literature emphasizing that valid higher-order-thinking instruments should represent the complexity of the construct being measured.

According to Sugiyono (2017), a research instrument functions as a systematic tool for collecting data and measuring the phenomena under investigation. In educational research, the quality of an instrument directly affects the credibility of conclusions regarding instructional effectiveness, student achievement, or learning-related change. Ridwan (2023) similarly emphasized that each assessment item should undergo systematic analysis before use to ensure that it is appropriate, understandable, and aligned with the intended construct. Validity and reliability are therefore essential because they determine whether an instrument measures what it intends to measure and produces sufficiently consistent evidence across respondents and administrations (Lahantaya, 2020; Saputra et al., 2022; Rodríguez-Rojas et al., 2024). Without adequate psychometric evidence, improvements in students' scores may be misinterpreted because they could reflect weak item construction, inconsistent scoring, or construct underrepresentation rather than genuine critical-thinking development.

Essay-based assessments are particularly relevant for measuring critical thinking because they enable students to construct explanations, justify positions, connect evidence with claims, and articulate their reasoning processes. Unlike selected-response tests, essay questions can reveal the quality of students' interpretation, argumentation, and evaluation when supported by transparent scoring rubrics. In science education, constructed-response tasks are especially valuable for assessing how students reason about phenomena that involve uncertainty, multiple causal factors, and context-dependent solutions. Recent studies have emphasized that critical-thinking assessment should include tasks that require students to explain why a conclusion is justified, evaluate available evidence, and formulate responses to authentic problems rather than simply identify correct answers from predetermined options (Bhakti et al., 2025; Rodríguez-Rojas et al., 2024; Sujanem & Suwindra, 2023). Accordingly, the present study uses essay-based pretest and posttest instruments designed to capture students' critical-thinking performance in the context of environmental-change learning.

The use of parallel pretest and posttest instruments is also methodologically important. When the same test items are administered repeatedly, students' scores may be influenced by memory effects, familiarity with question formats, or exposure to previously encountered content. Parallel forms seek to minimize these threats by using different items that represent equivalent constructs, blueprints, difficulty expectations, and scoring criteria. This approach is particularly relevant in intervention studies evaluating digital learning media because researchers need to distinguish genuine learning gains from test-retest familiarity. Research on technology-supported assessment has likewise underscored the importance of ensuring comparability, accessibility, construct alignment, and scoring quality when assessments are delivered across changing instructional contexts (Perry et al., 2022; Rodríguez-Rojas et al., 2024; Zen et al., 2022).

Previous studies have developed critical-thinking instruments in various science, mathematics, physics, and technology-related contexts. For example, Bhakti et al. (2025) developed and validated a complex true-false multiple-choice assessment for critical thinking in physics, while Rodríguez-Rojas et al. (2024) developed a broader critical-thinking evaluation scale within a Colombian context. Indonesian studies have also developed instruments and

learning media intended to foster higher-order thinking, including inquiry-based assessments, digital modules, and local-wisdom-oriented learning materials (Chusni et al., 2021; Isnaeni et al., 2021; Sumarni & Kadarwati, 2020). Although these studies provide valuable evidence, many focus on general instructional contexts, particular subject domains, or media effectiveness rather than the development of parallel essay-based instruments specifically designed for contextual digital learning media involving local environmental wisdom.

Moreover, prior studies on environmental-change learning have frequently concentrated on the effectiveness of inquiry, discovery learning, problem-based learning, or digital media in improving critical-thinking skills. Chusni et al. (2021), for instance, reported that discovery learning supported by e-learning could strengthen students' critical-thinking skills in environmental-change learning. Likewise, Suryawati et al. (2020) demonstrated the potential of local environmental problem-based worksheets to strengthen environmental literacy, while Amanda et al. (2025) developed an environmental-change e-module intended to support critical-thinking development. However, limited attention has been directed toward the psychometric development of equivalent pretest and posttest essay instruments that are explicitly aligned with local-wisdom-based digital media, project-based learning stages, and Facione's six critical-thinking indicators. This limitation is important because intervention effectiveness cannot be interpreted confidently without evidence that the instruments used to measure learning outcomes are valid, reliable, and contextually appropriate.

The present study addresses this gap by developing and validating parallel cognitive assessment instruments for Sc-SEBION-based learning on environmental change. The instruments are specifically designed to measure six dimensions of critical thinking interpretation, analysis, evaluation, inference, explanation, and self-regulation using contextual essay tasks related to *Syzygium cumini* conservation and environmental change. The study contributes to the literature in three important ways. First, it integrates a widely recognized critical-thinking framework with local ecological knowledge and culturally meaningful environmental issues. Second, it develops parallel pretest and posttest forms rather than relying on a single repeated instrument. Third, it provides preliminary content-validity, empirical-validity, and reliability evidence for an assessment instrument designed to accompany interactive digital learning media.

Accordingly, this study aims to develop and validate parallel essay-based cognitive assessment instruments for measuring junior high school students' critical-thinking skills in Sc-SEBION-based environmental-change learning. Specifically, the study seeks to: (1) determine the content validity of the developed pretest and posttest instruments; (2) examine the empirical validity of each item through item-total correlation analysis; and (3) determine the internal consistency reliability of both instrument forms. Through these objectives, the study is expected to provide a contextually relevant and psychometrically informed assessment foundation for evaluating the implementation and future effectiveness of Sc-SEBION learning media.

## METHODS

This study focused on the development and validation of cognitive assessment instruments intended to measure students' critical thinking skills in Sc-SEBION-based learning. The instruments consisted of parallel pretest and posttest assessments and were developed using a research and development (R&D) approach based on the 4D model, which comprises the define, design, develop, and disseminate stages (Thiagarajan et al., 1974). The 4D model was employed specifically to guide the systematic development, revision, and validation of the assessment instruments rather than the Sc-SEBION media itself. This approach was considered appropriate because rigorous instrument development requires a structured sequence beginning with construct identification and item generation, followed by expert review, pilot testing, and psychometric evaluation (Boateng et al., 2018; Stefana et al., 2025).

At the define stage, curriculum and learning outcomes, student characteristics, concepts, objectives, and task requirements were analyzed to identify learning needs and determine the

critical thinking indicators. This stage followed the analytical orientation of the 4D model, which emphasizes front-end analysis, learner analysis, task analysis, concept analysis, and specification of instructional objectives before product development begins (Thiagarajan et al., 1974). In addition, the construct definition process was informed by the principle that assessment instruments should be developed from a clearly delimited theoretical domain to ensure that the resulting items adequately represent the intended construct (Boateng et al., 2018; Romero Jeldres et al., 2023; Stefana et al., 2025).

At the design stage, the instrument blueprint was developed, including indicators and sub-indicators, learning objective flow, test-item indicators, question construction, cognitive domains, difficulty levels, item types, answer keys, and scoring rubrics. The cognitive assessment instruments consisted of pretest and posttest forms, each containing 12 essay questions based on Facione's six critical thinking indicators: interpretation, analysis, evaluation, inference, explanation, and self-regulation (Facione, 1990). The two forms were designed as parallel instruments with different question stimuli but equivalent constructs, blueprints, cognitive demands, and scoring weights. Essay-based items were selected because they allow students to articulate reasoning, interpret environmental phenomena, evaluate evidence, construct explanations, and justify conclusions rather than merely identify predetermined answers.

During the develop stage, the draft instruments underwent expert validation, revision, pilot testing, and statistical analysis. Content validity was evaluated by four validators, consisting of three university lecturers and one junior high school science teacher with expertise in science education, educational assessment, instructional design, and science learning. The use of expert judgment was intended to obtain evidence regarding the relevance, representativeness, clarity, and appropriateness of each item in relation to the targeted construct and learning context (Lawshe, 1975; Polit & Beck, 2006; Romero Jeldres et al., 2023). The validators reviewed the instruments in terms of content relevance, alignment with critical thinking indicators, item construction, narrative stimulus, language clarity, and instruction clarity.

Each item was assessed using a four-point Likert scale, ranging from 1 = very inappropriate, 2 = inappropriate, 3 = appropriate, to 4 = very appropriate. The use of a four-point scale was intended to encourage clear expert judgments by avoiding a neutral midpoint. The validation scores were subsequently converted into percentages to determine the preliminary level of expert agreement and content feasibility of each item. This descriptive percentage-based procedure was used as an initial indicator of item appropriateness, while expert suggestions were analyzed qualitatively to identify aspects requiring revision. Recommendations from validators were then used to refine the wording of the narrative stimuli, clarify question prompts, strengthen the alignment between items and critical thinking indicators, and improve scoring rubrics before empirical testing. This iterative revision process is consistent with recommended practices in instrument development, which emphasize the integration of expert feedback before pilot administration and psychometric evaluation (Boateng et al., 2018; Stefana et al., 2025). The criteria used to classify the level of content validity are presented in the following table.

**Table 1.** Criteria for Content Validity of the Cognitive Assessment Instrument

| No. | Percentage (%) | Validity Category |
|---|---|---|
| 1 | $p > 85$ | Very Valid |
| 2 | $75 < p \leq 85$ | Valid |
| 3 | $40 < p \leq 75$ | Fairly valid |
| 4 | $20 < p \leq 40$ | Less Valid |
| 5 | $p \leq 20$ | Invalid |

*Note: (p) = percentage score obtained from the content validity assessment results.*

After the content validity assessment was completed, the instrument underwent empirical validity testing through a pilot study involving 21 students who were not included in the main research sample. The pilot test was conducted with students in Class VIII-B. This class was selected as a convenience pilot group with prior exposure to Grade VII environmental

change content, which ensured that respondents had sufficient baseline understanding to engage with the test items meaningfully during initial item analysis. To minimize potential bias arising from prior knowledge, the pilot testing was not used to measure learning achievement, but solely to evaluate item quality, including item discrimination and internal consistency.

Empirical validity was analyzed by examining the correlation between individual item scores and the total test score using the Pearson item–total correlation in SPSS version 29. The validity of each item was determined by comparing the obtained correlation coefficient (r-count) with the critical r-table value at the 5% significance level. Items with r-count values greater than the r-table value were considered empirically valid. Furthermore, the magnitude of the correlation coefficients was interpreted according to the validity classification criteria presented in Table 2, ranging from very low validity to very high validity. Instrument reliability was analyzed using Cronbach's Alpha coefficient to determine the internal consistency of the instrument. The resulting reliability coefficient was interpreted according to the reliability classification criteria presented in Table 3, ranging from poor reliability to very high reliability. These interpretation criteria were used to describe the quality of the validity and reliability coefficients obtained from the statistical analyses.

**Table 2.** Interpretation of Instrument Validity

| Correlation Coefficient (rxy) r-count value | Interpretation |
|---|---|
| 0,80 < r xy ≤ 1,00 | Very High Validity |
| 0,60 < r xy ≤ 0,80 | High Validity |
| 0,40 < r xy ≤ 0,60 | Moderate Validity |
| 0,20 < r xy ≤ 0,40 | Low Validity |
| 0,00 < r xy ≤ 0,20 | Very Low Validity |
| r xy ≤ 0,00 | Invalid |

*Note: (r xy) = the correlation coefficient between item scores (X) and total test scores (Y) obtained from the Pearson item–total correlation analysis.*

**Table 3.** Interpretation of Instrument Reliability

| Cronbach's Alpha | Reliability Level |
|---|---|
| 0,00 ≤ r < 0,20 | Poor Reliability |
| 0,20 ≤ r < 0,40 | Fair Reliability |
| 0,40 ≤ r < 0,60 | Moderate Reliability |
| 0,60 ≤ r < 0,80 | Good Reliability |
| 0,80 ≤ r < 1,00 | Very High Reliability |

*Note: (r) = the reliability coefficient obtained from Cronbach's Alpha analysis.*

The overall procedure for developing the pretest and posttest is presented in Figure 1.

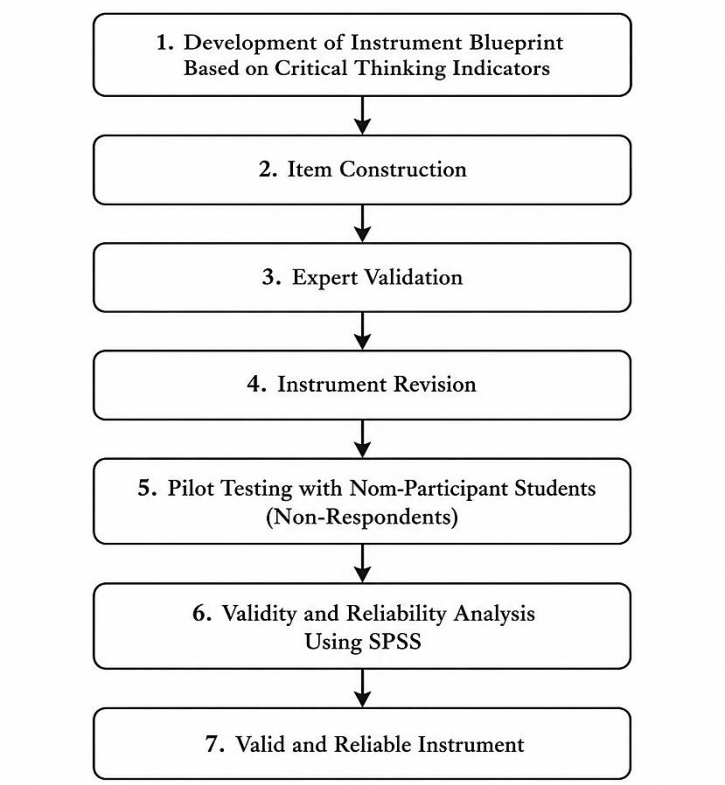


**Figure 1**. Research Procedure Flowchart

The final stage of the 4D model is the Disseminate stage. In this study, dissemination was restricted to the preparation and readiness of the validated instruments for application in Sc-SEBION-based learning activities within the study context.

## RESULT AND DISCUSSION

### *Result*

Sc-SEBION is a digital learning media based on the explosion box concept that is designed with multiple interactive layers to support contextual and engaging learning experiences. The media adopts the working principle of an explosion box, which literally refers to a "box that unfolds." When accessed, each layer of the media gradually reveals different learning content (Fatmawati *et al.*, 2024). These layers are organized according to the stages of Project-Based Learning (PjBL), providing students with a structured learning experience. The content includes information, visual representations, and learning activities related to science concepts, particularly environmental change topics (Khoiriyah *et al.*, 2025).

The topic of environmental change was selected because the population of juwet plants (*Syzygium cumini*) has recently experienced a significant decline. This decline is not only caused by human activities, such as tree cutting and land conversion, but is also influenced by climate change that triggers broader environmental changes. Increasing temperatures, extreme weather conditions, and ecosystem disturbances can negatively affect the growth, regeneration, and sustainability of juwet plants. As a result, environmental change became a relevant and contextual topic to be integrated into the Sc-SEBION learning media in order to increase students' awareness of local environmental problems and conservation efforts.

Conceptually, an explosion box is a three-dimensional (3D) learning media designed to stimulate students' curiosity, imagination, and critical thinking through the gradual exploration of its contents (Safitri *et al.*, 2024; Nasriyah *et al.*, 2021). In Sc-SEBION, this concept has been transformed into an interactive digital format by incorporating 3D visualizations, animations, and exploratory features such as zoom functions and pop-up content. These features encourage active student engagement throughout the learning process and have the potential to improve

learning outcomes (Fatmawati *et al*., 2024). To determine whether students' learning outcomes and critical thinking skills improve after using Sc-SEBION, appropriate cognitive assessment instruments are required. This consists of pretest and posttest instruments.

The pretest and posttest instruments were developed using the 4D model, which consists of four stages: Define, Design, Develop, and Disseminate. The Define stage established the assessment framework, including the topic of Environmental Change, the learning objectives, the cognitive skills to be assessed, the competency indicators, and the test blueprint. In this study, the development process was systematically aligned with the curriculum structure, beginning with the learning topic (Environmental Change) and the Learning Outcomes (LOs), which state that students are expected to analyze interactions between living organisms and their environment in efforts to mitigate climate change. Based on these LOs, the learning objectives were operationalized into specific learning progressions, including identifying environmental phenomena, analyzing causes of environmental change, and evaluating its ecological impacts. Subsequently, indicators and sub-indicators were derived from the learning progression and translated into item indicators aligned with Facione's critical thinking framework.

The Design stage resulted in the initial drafts of the cognitive assessment instruments, including essay items and scoring rubrics. Each item was constructed based on the systematic alignment among the Learning Outcomes (LOs), learning progression, indicators, sub-indicators, and item indicators, ensuring strong content and construct validity. The Develop stage focused on validating and refining the instruments through expert judgment and iterative revision to ensure validity and reliability. As a result, the revised instruments met the established criteria and were deemed suitable for implementation.

The final instruments consisted of two parallel forms, namely pretest and posttest, with different items but identical constructs, blueprint, and weighting, indicating equivalence in content and structure. Both pretest and posttest comprised 12 essay items and were designed to measure six critical thinking skills proposed by Facione: interpretation, analysis, evaluation, inference, explanation, and self-regulation. In addition, the items were integrated within the stages of Project-Based Learning (PjBL), including problem identification, environmental issue analysis, solution development, collaborative project implementation, strategy evaluation, and conclusion drawing related to juwet (*Syzygium cumini*) conservation. This integration ensured that the instruments assessed not only conceptual understanding but also higher-order thinking skills in a contextual and authentic learning environment.

To illustrate the alignment among the learning topic, Learning Outcomes, learning progression, indicators, and item indicators, sample items from both instruments are presented in table 4 and 5 below.

**Table 4.** Pretest Example (Interpretation – identifying current environmental phenomena)

| **Indicator and Sub-Indicator** | **Interpretation: interpreting the meaning of phenomena** |
|---|---|
| Learning Progression | Interpreting environmental change concepts based on presented phenomena |
| Item Indicator | Students interpret the impact of climate change on juwet (*Syzygium cumini*) population |
| Question | Read the following passage!<br>Lopang Village was previously covered by a large number of juwet (*Syzygium cumini*) trees, which are part of local cultural heritage. However, in recent years, the population of these trees has declined due to climate change. Question:<br>Interpret this phenomenon in relation to climate change! |
| Cognitive Domain | C4 |

| | |
|---|---|
| Item Type | Essay |
| Answer Key | Climate change leads to environmental changes that are not conducive to the growth and regeneration of *Syzygium cumini*, resulting in a decline in its population. (key concepts: climate change → environmental change → unsuitable growth conditions → population decline) |
| Scoring Rubric | 4= all four key concepts with correct explanation<br>3= three correct key concepts<br>2= two correct key concepts<br>1 = one correct key concept<br>0 = no answer / all incorrect |

**Table 5.** Posttest Example (Interpretation – analyzing future environmental consequences)

| **Indicator and Sub-Indicator** | **Interpretation: interpreting the meaning of phenomena** |
|---|---|
| Learning Progression | Interpreting environmental change concepts based on presented phenomena |
| Item Indicator | Students interpret statements regarding the future condition of juwet (*Syzygium cumini*) under continued climate change |
| Question | Read the following statement:<br>If environmental conditions continue to change as they currently do, *Syzygium cumini* (juwet) trees will become increasingly difficult to find in the future.<br>Question:<br>Interpret in your own words the meaning of this statement regarding the future of *Syzygium cumini* if climate change continues. |
| Cognitive Domain | C4 |
| Item Type | Essay |
| Answer Key | Climate change leads to environmental instability, which results in conditions that are not conducive to the growth and regeneration of *Syzygium cumini*. As a consequence, trees become less able to survive and reproduce, leading to a decline in population and reduced availability in the future. |
| Scoring Rubric | 4= more than four correct key concepts with accurate explanation<br>3= three correct key concepts<br>2= two correct key concepts<br>1= one correct key concept<br>0= no answer / all incorrect |

These examples demonstrate that each item was systematically derived from the alignment of the Learning Outcomes, learning progression, indicators, sub-indicators, and item indicators, ensuring construct relevance and content validity. The results of the content validity assessment are presented in Tables 6 and 7.

**Table 6.** Expert Validation Results of the Pretest Cognitive Assessment Instrument

| **No.** | **Assessment Criteria** | **Mean Score** | **Category** |
|---|---|---|---|
| 1 | Alignment with learning objectives | 3.67 | Very Valid |
| 2 | Alignment with material concepts | 3.85 | Very Valid |

| | | | |
|---|---|---|---|
| 3 | Alignment with critical thinking indicators | 3.69 | Very Valid |
| 4 | Clear and understandable wording | 3.77 | Very Valid |
| 5 | Appropriate question construction | 3.83 | Very Valid |
| 6 | Clear instructions | 3.83 | Very Valid |
| 7 | Relevant narrative stimulus | 3.65 | Very Valid |
| 8 | Clear and communicative language | 3.85 | Very Valid |
| 9 | Unambiguous sentences | 3.82 | Very Valid |
| | Overall Mean Score | 3.77 | Very Valid |

*Note. Scores were rated on a four-point scale. P = question item number.*

Based on the expert validation results of the pretest instrument, the validity percentages for the 12 test items (P1–P12) ranged from 93.1% to 96.5%. All items were classified as Very Valid according to the established criteria. Item P1 obtained the highest validity score (96.5%), while items P8 and P12 received the lowest scores (93.1%). Although all items met the criteria for use, the validators provided several suggestions for improvement. In particular, the slightly lower scores for P8 and P12 were related to the clarity of the narrative stimulus and the alignment of the questions with the intended critical thinking indicators. Based on these comments, revisions were made to improve the wording of the stimulus, clarify the question prompts, and strengthen the correspondence between the item content and the targeted indicators. These revisions were implemented before the empirical testing stage to enhance the overall quality of the instrument.

**Table 7.** Expert Validation Results of the Posttest Cognitive Assessment Instrument

| No. | Assessment Criteria | Mean Score | Category |
|---|---|---|---|
| 1 | Alignment with learning objectives | 3.6 | Very Valid |
| 2 | Alignment with material concepts | 3.79 | Very Valid |
| | Overall Mean Score | 3.7 | Very Valid |

*Note. Scores were rated on a four-point scale.*

Based on the posttest instrument validation results, the validity percentages for the 12 test items (P1–P12) ranged from 91.7% to 96.5%. All items were categorized as Very Valid, indicating that the instrument met the required criteria in terms of content, construction, and language. Item P1 achieved the highest validity percentage (96.5%), followed by P10 (95.8%) and P11 (95.1%). Meanwhile, items P3, P4, P5, P7, P8, and P12 obtained a validity percentage of 91.7%. Although these values were slightly lower than those of the other items, they still fell within the Very Valid category. The high validity scores obtained by both the pretest and posttest instruments indicate that the developed items are capable of accurately measuring students' cognitive abilities in accordance with the intended learning objectives. Furthermore, the high scores achieved in the aspect of alignment with critical thinking indicators suggest that the instruments assess not only students' conceptual understanding but also their higher-order thinking skills. Although several items received relatively lower scores on the narrative stimulus aspect, all items remained within the Very Valid category and therefore did not require substantial revisions.

The average validity scores of the pretest and posttest instruments differed by only 0.80%, indicating that both instruments possess comparable quality. Consequently, the developed pretest and posttest instruments were considered suitable for measuring students' cognitive abilities before and after instruction, thereby providing valid data for evaluating the effectiveness of the implemented learning process. Following the confirmation of content validity, empirical validity testing was conducted. Responses from 21 non-participant students were analyzed using Pearson item–total correlation, and the results are presented in Tables 8.

**Table 8.** The Result of Analysis Pearson-Total Correlation of Pretest and Posttest by Using SPSS 29

| Correlations | Pretest | Posttest |
|---|---|---|

| | | Total | Total |
|---|---|---|---|
| P1 | Pearson Correlation | .606** | .790** |
| | Sig. (2-tailed) | .004 | <.001 |
| | N | 21 | 21 |
| P2 | Pearson Correlation | .848** | .769** |
| | Sig. (2-tailed) | <.001 | <.001 |
| | N | 21 | 21 |
| P3 | Pearson Correlation | .742** | .695** |
| | Sig. (2-tailed) | <.001 | <.001 |
| | N | 21 | 21 |
| P4 | Pearson Correlation | .807** | .733** |
| | Sig. (2-tailed) | <.001 | <.001 |
| | N | 21 | 21 |
| P5 | Pearson Correlation | .750** | .823** |
| | Sig. (2-tailed) | <.001 | <.001 |
| | N | 21 | 21 |
| P6 | Pearson Correlation | .840** | .916** |
| | Sig. (2-tailed) | <.001 | <.001 |
| | N | 21 | 21 |
| P7 | Pearson Correlation | .927** | .657** |
| | Sig. (2-tailed) | <.001 | .001 |
| | N | 21 | 21 |
| P8 | Pearson Correlation | .781** | .797** |
| | Sig. (2-tailed) | <.001 | <.001 |
| | N | 21 | 21 |
| P9 | Pearson Correlation | .696** | .718** |
| | Sig. (2-tailed) | <.001 | <.001 |
| | N | 21 | 21 |
| P10 | Pearson Correlation | .789** | .810** |
| | Sig. (2-tailed) | <.001 | <.001 |
| | N | 21 | 21 |
| P11 | Pearson Correlation | .805** | .744** |
| | Sig. (2-tailed) | <.001 | <.001 |
| | N | 21 | 21 |
| P12 | Pearson Correlation | .714** | .802** |
| | Sig. (2-tailed) | <.001 | <.001 |
| | N | 21 | 21 |
| Total | Pearson Correlation | 1 | 1 |
| | Sig. (2-tailed) | | |
| | N | 21 | 21 |

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 8 shows that the calculated correlation coefficients (r-count) for all pretest and posttest items were higher than the critical r-table value of 0.433. In addition, all items yielded significance values (Sig. 2-tailed) below 0.05, with most items showing significance levels of < 0.001. The empirical validity test involved responses from 21 students who were not included as research participants. The high correlation coefficients indicate a strong positive relationship between each item score and the total test score. This finding suggests that each item contributes effectively to measuring the same construct represented by the overall instrument. Furthermore, the significance values lower than 0.05 indicate that the correlations are statistically significant, confirming the empirical validity of the test items. Therefore, all pretest and posttest items met the validity criteria and were considered appropriate for use in measuring students' critical thinking skills. The r-count for each item in both the pretest and posttest instruments are summarized in Table 9.

**Table 9.** Empirical Validity of the Pretest and Posttest Items

| Validity of the Cognitive Assessment Instruments | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Pretest | | | Posttest | | |
| Item | r-table | r-count | Validity | Validity Category | r-count | Validity | Validity Category |
| P1 | 0.433 | 0,606 | Valid | High Validity | 0,790 | Valid | High Validity |
| P2 | 0.433 | 0,848 | Valid | Very High Validity | 0,769 | Valid | High Validity |
| P3 | 0.433 | 0,742 | Valid | High Validity | 0,695 | Valid | High Validity |
| P4 | 0.433 | 0,807 | Valid | Very High Validity | 0,733 | Valid | High Validity |
| P5 | 0.433 | 0,750 | Valid | High Validity | 0,823 | Valid | Very High Validity |
| P6 | 0.433 | 0,840 | Valid | Very High Validity | 0,916 | Valid | Very High Validity |
| P7 | 0.433 | 0,927 | Valid | Very High Validity | 0,657 | Valid | High Validity |
| P8 | 0.433 | 0,781 | Valid | High Validity | 0,797 | Valid | High Validity |
| P9 | 0.433 | 0,696 | Valid | High Validity | 0,718 | Valid | High Validity |
| P10 | 0.433 | 0,789 | Valid | High Validity | 0,810 | Valid | Very High Validity |
| P11 | 0.433 | 0,805 | Valid | Very High Validity | 0,744 | Valid | High Validity |
| P12 | 0.433 | 0,714 | Valid | High Validity | 0,802 | Valid | Very High Validity |

*Note: (P) = question item number*

Table 9 presents a comparison between the calculated correlation coefficients (r-count) and the critical r-table value for each pretest and posttest item. In this study, the critical r-table value was 0.433. The analysis showed that all items in both the pretest and posttest instruments had r-count greater than the r-table value (0.433). Therefore, all items (P1–P12) were considered valid and suitable for measuring students' cognitive abilities. For the pretest instrument, the item–total correlation coefficients ranged from 0.606 to 0.927. Similarly, the posttest instrument showed item–total correlation coefficients ranging from 0.657 to 0.916. Since all correlation coefficients exceeded the critical r-table value, every item met the empirical validity criteria. Based on Sugiyono's validity classification, most items fell within the High Validity (0.60–0.799) and Very High Validity (0.80–1.00) categories. In the pretest instrument, five items (P2, P4, P6, P7, and P11) were classified as having Very High Validity. Likewise, five items in the posttest instrument (P5, P6, P10, P11, and P12) were categorized as having Very High Validity. These findings indicate that the majority of the items have a strong association with the construct being measured and are highly suitable for assessing students' critical thinking skills. In addition to validity, an instrument must also demonstrate a high level of reliability before it can be used in research. Reliability analysis was therefore conducted to evaluate the consistency of the pretest and posttest instruments. The results of the reliability analysis are presented in Tables 10.

**Table 10.** Reliability Results of the Pretest and Posttest Instrument

| Reliability Statistics | | | |
|---|---|---|---|
| Pretest | | Posttest | |
| Cronbach's Alpha | N of Items | Cronbach's Alpha | N of Items |
| .936 | 12 | .934 | 12 |

Based on the reliability analysis presented in Table 10 (left side), the pretest instrument obtained a Cronbach's Alpha coefficient of 0.936 across 12 items. Reliability refers to the extent to which an instrument produces consistent results when used to measure the same construct under similar conditions. According to Sugiyono's reliability criteria, a Cronbach's Alpha value within the range of $0.80 \leq r < 1.00$ is categorized as Very High Reliability. Therefore, the pretest instrument demonstrated a very high level of internal consistency, indicating that the items consistently measure the intended construct of students' critical thinking skills.

Table 10 in the right side presents the results of Cronbach's Alpha reliability analysis for the posttest instrument. The analysis yielded a Cronbach's Alpha coefficient of 0.934 for the 12-item instrument. Based on Sugiyono's reliability criteria, this value falls within the range of 0.80

≤ r < 1.00 and is therefore categorized as Very High Reliability. This result indicates that the posttest instrument has a very high level of internal consistency and is appropriate for measuring students' critical thinking skills after the implementation of the Sc-SEBION learning media.

The very high reliability coefficient suggests that all items in the instrument consistently measure the same construct. This finding indicates that the posttest instrument is capable of producing stable and dependable data. Overall, the reliability analysis of both the pretest (α = 0.936) and posttest (α = 0.934) instruments demonstrates that they possess very high internal consistency. Although Cronbach's Alpha values above 0.90 may indicate strong internal consistency, they may also raise concerns regarding potential item redundancy. However, in this study, the instrument was intentionally designed to measure multiple dimensions of critical thinking based on Facione's framework. Specifically, items 1–2 measure interpretation, items 3–4 measure analysis, items 5–6 measure evaluation, items 7–8 measure inference, items 9–10 measure explanation, and items 11–12 measure self-regulation. Therefore, the high reliability coefficient does not necessarily indicate redundancy, but rather reflects consistency among items within and across the defined critical thinking indicators. Nevertheless, the present study did not conduct construct validation (e.g., exploratory or confirmatory factor analysis) to statistically confirm the dimensional structure of the instrument. Therefore, further studies are recommended to examine whether the theoretical six-factor structure of Facione's critical thinking framework is empirically supported and to provide stronger evidence of construct validity. The last stage is the dissemination stage that was restricted to the preparation and readiness of the validated instruments for application in Sc-SEBION-based learning activities within the study context.

### *Discussion*

The present study developed parallel pretest and posttest essay instruments to assess junior high school students' critical thinking skills within Sc-SEBION-based environmental change learning. The findings demonstrate that both instruments achieved strong preliminary psychometric quality. Content validation results ranged from 93.1% to 96.5% for the pretest and from 91.7% to 96.5% for the posttest, indicating that the items were highly aligned with the learning objectives, environmental-change content, critical-thinking indicators, item construction, and language requirements. Furthermore, all 12 items demonstrated empirical validity, with item-total correlations ranging from .606 to .927 for the pretest and from .657 to .916 for the posttest. The internal consistency coefficients were also very high for both forms, with Cronbach's alpha values of .936 and .934, respectively. Importantly, the small difference in average content-validity scores between the two forms suggests that the pretest and posttest possess comparable quality for assessing students before and after Sc-SEBION-based instruction.

These results are consistent with international studies emphasizing that critical-thinking assessment must be grounded in explicit constructs, contextual learning objectives, and coherent scoring procedures. Suprapto et al. (2020) similarly reported that higher-order-thinking instruments require systematic alignment between indicators, learning content, and assessment tasks. Likewise, Abeden and Siew (2022) demonstrated that critical-thinking assessment in science-related disciplines should capture students' capacity to analyse information, evaluate evidence, and formulate reasoned conclusions rather than merely recall factual knowledge. The strong item-total correlations obtained in the present study therefore indicate that each essay item contributes meaningfully to the broader construct of critical thinking.

The present findings also correspond with the development of more rigorous critical-thinking measures internationally. Rodríguez-Rojas et al. (2024) showed that the development of a critical-thinking scale requires evidence beyond conceptual relevance, including reliability and construct-related validation. Similarly, Bhakti et al. (2025) highlighted the importance of developing assessment tasks that can capture higher-order reasoning rather than relying exclusively on conventional factual multiple-choice tests. In this regard, the essay format used in

the current study is pedagogically appropriate because it allows students to explain causal relationships, interpret environmental phenomena, justify judgments, and propose solutions related to the decline of *Syzygium cumini*. Such responses are more compatible with the interpretative, analytical, evaluative, inferential, explanatory, and self-regulatory dimensions proposed in Facione's framework.

Nevertheless, the exceptionally high Cronbach's alpha values should be interpreted cautiously. Although coefficients above .90 indicate excellent internal consistency, they may also suggest that some items share highly similar cognitive demands. The parallel forms in this study were intentionally constructed to assess the same six dimensions of critical thinking using different environmental narratives. Thus, the high alpha values may reflect consistent construct representation rather than unnecessary item repetition. However, future analyses should examine item redundancy, dimensionality, and local item dependence through exploratory factor analysis, confirmatory factor analysis, or Rasch modelling. Recent Indonesian studies have increasingly used Rasch analysis to evaluate item fit, person reliability, rating-scale functioning, and construct structure in critical-thinking assessment instruments (Juandi et al., 2024; Noperi et al., 2025).

The study also contributes to global discussions concerning digital and project-based science education. Previous studies have shown that project-based and technology-supported learning environments can create opportunities for inquiry, evidence evaluation, collaboration, and reflective problem-solving. Almulla (2020) and Guo et al. (2020) argued that project-based learning promotes active participation when students are required to investigate meaningful problems and produce evidence-based outputs. In online environments, Cortázar et al. (2021) found that structured project-based activities can promote critical thinking when digital tasks are supported by clear scaffolding, feedback, and collaborative dialogue. Comparable findings were reported by Zen et al. (2022), who demonstrated that online project-based learning can strengthen engagement and academic outcomes when digital activities are deliberately designed rather than merely transferred from conventional classroom procedures.

The relevance of this literature to the current study lies not in claiming that Sc-SEBION has already improved students' critical-thinking skills, because the present research focused on instrument development rather than intervention effectiveness. Instead, the validated instruments provide a necessary measurement foundation for future effectiveness testing. This distinction is essential. A valid and reliable instrument enables researchers to determine whether digital media, contextual instruction, or project-based activities genuinely improve critical thinking rather than simply produce higher test scores through familiarity with content or repeated exposure to assessment tasks. Perry et al. (2022) similarly noted that digital assessment should be designed carefully to preserve construct validity when assessment formats shift from conventional to technology-supported environments.

In the Indonesian context, the results align with previous studies showing that contextual, digital, and inquiry-oriented materials are relevant for strengthening critical thinking. Chusni et al. (2021) found that discovery learning through e-learning on environmental-change topics encouraged students to engage more actively with information, although critical-thinking achievement remained uneven across learners. Primasari et al. (2020) also reported a meaningful relationship between science literacy, critical thinking, and learning motivation. These findings support the use of environmental-change scenarios in the present instrument because such scenarios allow students to connect scientific concepts with ecological problems that are visible in their local surroundings.

The integration of local wisdom through the *juwet* conservation context also represents an important contribution. Rather than treating environmental change as an abstract global issue, the instrument asks students to interpret ecological disruption through a culturally meaningful and geographically familiar case. This approach is consistent with Indonesian studies showing that contextual learning materials and local environmental issues can strengthen students' engagement with scientific phenomena (Martawijaya et al., 2023; Putri & Asrizal, 2023).

Similarly, Marlina et al. (2022) and Yunita et al. (2024) demonstrated that digital teaching materials designed around critical-thinking indicators can support more active learning when content, media, and assessment are systematically integrated. The present study extends this approach by developing not only interactive media but also parallel assessment instruments specifically mapped to six dimensions of critical thinking.

Theoretical, pedagogical, and policy implications should be considered together. Theoretically, the findings provide preliminary support for Facione's critical-thinking framework because the instrument blueprint was successfully operationalised into interpretation, analysis, evaluation, inference, explanation, and self-regulation items. The strong expert-validation scores suggest that these dimensions can be translated into contextual essay tasks involving environmental change and local biodiversity. At the same time, the findings reinforce the view that digital competence and critical thinking are related but conceptually distinct. Digital media alone does not automatically produce critical reasoning; rather, students require tasks that direct them to evaluate information, compare evidence, identify causal mechanisms, and justify conclusions (van Laar et al., 2020; Amin et al., 2023; Samura & Darhim, 2023).

Pedagogically, teachers may use the instruments as formative and summative tools to identify which critical-thinking dimensions require further support. For example, students who struggle with interpretation may need clearer environmental stimuli, whereas students who demonstrate weak evaluation skills may benefit from tasks requiring evidence comparison or the appraisal of alternative conservation strategies. At the policy level, schools should support the development of context-sensitive digital assessment systems, teacher assessment literacy, and ethical procedures for digital data management. However, the present study did not measure students' emotional well-being, online identity formation, or mental-health outcomes. Therefore, policy claims in these areas should remain limited. Future digital-learning policies should prioritise responsible technology use, equitable device access, assessment fairness, student-data protection, and teacher capacity to interpret assessment results meaningfully (Faradillah & Adlina, 2021; Yanuarto et al., 2021; Lestari et al., 2025).

The novelty of this study lies in the development of two parallel essay-based critical-thinking instruments specifically designed for Sc-SEBION learning on environmental change. Unlike generic critical-thinking tests, the instruments integrate Facione's six indicators with local ecological knowledge concerning *Syzygium cumini*, Project-Based Learning stages, and digital interactive-media implementation. This design contributes to the growing literature on contextual digital assessment by demonstrating that critical-thinking measurement can be connected with local biodiversity, environmental conservation, and culturally relevant science learning. Recent mapping studies have stressed the need for more contextual, psychometrically tested, and educationally meaningful critical-thinking assessments in STEM-related learning environments (Mudi et al., 2026; Wulandari et al., 2025).

Several limitations should be acknowledged. First, empirical validation involved only 21 non-participant students, limiting the stability and generalisability of the item-total correlations and reliability coefficients. Second, the study employed classical test theory and did not conduct factor analysis, Rasch analysis, test-information analysis, measurement invariance testing, or criterion-related validation. Third, because the instrument used essay tasks, inter-rater reliability should be examined in future studies to ensure scoring consistency across teachers or evaluators. Fourth, the study did not test the effectiveness of Sc-SEBION media itself; therefore, no causal conclusion can be drawn regarding improvements in students' critical-thinking skills. Future research should involve larger and more diverse samples, multiple schools, quasi-experimental or experimental designs, Rasch-based psychometric testing, inter-rater agreement analysis, and comparisons between Sc-SEBION-based instruction and conventional science learning. Such studies would provide stronger evidence regarding both the measurement quality of the instrument and the educational effectiveness of Sc-SEBION learning media.

## CONCLUSION

The results of this study indicate that the cognitive assessment instruments (pretest and posttest) developed to measure students' critical thinking skills in Sc-SEBION (*Syzygium cumini*–Explosion Box Interactive Online Version) learning demonstrate satisfactory levels of validity and reliability. Content validity analysis showed that all pretest and posttest items obtained validity percentages within the very valid category, indicating that the items were appropriate in terms of content and aligned with the intended indicators. Furthermore, empirical validity testing revealed that all 12 items had item-total correlation coefficients higher than the r-table value (0.433), confirming that all items were statistically valid and ranged from high to very high validity levels based on Sugiyono's criteria. These findings indicate that the items were consistent in measuring the intended construct of critical thinking skills. In addition, the reliability analysis showed Cronbach's Alpha values of 0.936 for the pretest and 0.934 for the posttest instruments, both of which fall within the very high category. These values indicate strong internal consistency, suggesting that the instruments produce stable and consistent measurement outcomes within the pilot sample. Overall, the findings suggest that the developed instruments are valid and reliable for measuring students' critical thinking skills in Sc-SEBION-based learning. However, given that the empirical testing was conducted on a relatively small sample (n = 21), the generalizability of the results remains limited. Therefore, the instruments should be interpreted as promising tools within the specific context of this study rather than fully standardized instruments. Further research involving larger and more diverse samples is recommended to strengthen evidence of validity, reliability, and generalizability.

## REFERENCE


Abeden, N. A. M., & Siew, N. M. (2022). Assessing students' critical thinking and physics problem-solving skills in secondary schools. *Malaysian Journal of Social Sciences and Humanities, 7*(6), e001584. https://doi.org/10.47405/mjssh.v7i6.1584

Almulla, M. A. (2020). The effectiveness of the project-based learning approach as a way to engage students in learning. *SAGE Open, 10*(3), Article 2158244020938702. https://doi.org/10.1177/2158244020938702

Amanda, D., Fajrina, S., Razak, A., & Darussyamsu, R. (2025). Development of a SETS-based e-module to enhance critical thinking skills on environmental change. *JPBI (Jurnal Pendidikan Biologi Indonesia), 11*(3), 859–873. https://doi.org/10.22219/jpbi.v11i3.41221

Amin, A. M., Adiansyah, R., & Hujjatusnaini, N. (2023). The contribution of communication and digital literacy skills to critical thinking. *Jurnal Pendidikan Sains Indonesia, 11*(3), 697–712. https://doi.org/10.24815/jpsi.v11i3.30838

Askhar, B. M., Megawati, M., Mahmud, M. B., & Rachman, F. N. (2022). Wisata petik buah sebagai habitus sosial ekonomi baru bagi masyarakat Dusun Sumberejo, Desa Lopang. *Sewagati: Jurnal Pengabdian Masyarakat Indonesia, 1*(2), 12–24. https://doi.org/10.56910/sewagati.v1i2.471

Bhakti, Y. B., Arthur, R., & Supriyati, Y. (2023). Development of an assessment instrument for critical thinking skills in physics: A systematic literature review. *Journal of Physics: Conference Series, 2596*(1), Article 012067. https://doi.org/10.1088/1742-6596/2596/1/012067

Bhakti, Y. B., Arthur, R., & Supriyati, Y. (2025). Development and validation of a complex true-false multiple-choice test for assessing critical thinking skills in physics. *Cogent Education, 12*(1), Article 2553008. https://doi.org/10.1080/2331186X.2025.2553008

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Chusni, M. M., Saputro, S., Rahardjo, S. B., & Suranto. (2021). Students' critical thinking skills through discovery learning model using e-learning on environmental change subject matter. *European Journal of Educational Research, 10*(3), 1123–1135. https://doi.org/10.12973/eu-jer.10.3.1123

Cortázar, C., Nussbaum, M., Harcha, J., Alvares, D., López, F., Goñi, J., & Cabezas, V. (2021). Promoting critical thinking in an online, project-based course. *Computers in Human Behavior, 119*, Article 106705. https://doi.org/10.1016/j.chb.2021.106705

Dwiprabowo, R., Herianingtyas, N. L. R., Zulfikri, T., Suryaningsih, T., Agustin, N. D., Azhar, A. P. S., Mahmud, M. B., Raihan, M. D., Setiahati, I. P., Fajri, H. M., Dewi, N., Sesrita, A., Sinyanyuri, S., Novianti, N., Sutisnawati, A., & Friska, S. Y. (2024). *Berpikir kritis dan berpikir kreatif siswa sekolah dasar*. Publica Indonesia Utama.

Ennis, R. H. (2011). *The nature of critical thinking: An outline of critical thinking dispositions and abilities*. University of Illinois. https://education.illinois.edu/docs/default-source/faculty-documents/robert-ennis/the-nature-of-critical-thinking_51711_000.pdf

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction: The Delphi report*. American Philosophical Association. https://eric.ed.gov/?id=ED315423

Faradillah, A., & Adlina, S. (2021). Validity of critical thinking skills instrument on prospective mathematics teachers. *Jurnal Penelitian dan Evaluasi Pendidikan, 25*(2), 126–137. https://doi.org/10.21831/pep.v25i2.40662

Fatmawati, F., Muhaemin, M., & Hasriadi, H. (2024). Development of explosion box learning media in Islamic Religious Education subject for students of junior high school. *Educational Journal of Learning Technology, 2*(1), 75–83. https://doi.org/10.58230/edutech.v2i1.32

Firmansyah, F. A. A. H., Subroto, W. T., & Hendratno. (2023). Development of argumentation-based 3D audio visual media exploding box to improve students' critical thinking in elementary schools. *Jurnal Pendidikan, Bahasa dan Budaya, 2*(2), 122–147. https://doi.org/10.55606/jpbb.v2i2.1510

Guo, P., Saab, N., Post, L. S., & Admiraal, W. (2020). A review of project-based learning in higher education: Student outcomes and measures. *International Journal of Educational Research, 102*, Article 101586. https://doi.org/10.1016/j.ijer.2020.101586

Haatainen, O., & Aksela, M. (2021). Project-based learning in integrated science education: Active teachers' perceptions and practices. *LUMAT: International Journal on Math, Science and Technology Education, 9*(1), 149–173. https://doi.org/10.31129/LUMAT.9.1.1392

Isnaeni, W., Sujatmiko, Y. A., & Pujiasih, P. (2021). Analysis of the role of Android-based learning media in learning critical thinking skills and scientific attitude. *Jurnal Pendidikan IPA Indonesia, 10*(4), 607–617. https://doi.org/10.15294/jpii.v10i4.27597

Juandi, T., Kaniawati, I., Samsudin, A., Riza, L., Susilawati, S., & Sapiruddin, S. (2024). Analysis of the validity and reliability of a critical thinking skills instrument on the topic of wave-particle duality using Rasch model. *Quantum: Jurnal Inovasi Pendidikan Sains, 15*(2), 197–209. https://doi.org/10.20527/quantum.v15i2.19996

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lestari, T., Rosidin, U., Distrik, I. W., Ertikanto, C., & Herlina, K. (2025). Analysis of assessment instrument needs to measure students' critical thinking skills and self-efficacy in physics learning. *Jurnal Penelitian Pendidikan IPA, 10*(12), 1–6. https://doi.org/10.29303/jppipa.v10i12.9453

Marlina, L., Paramitha, G. P., & Sriyanti, I. (2022). Development of electronic modules based on critical thinking skills on vibration, waves, and sound materials for junior high school students. *Jurnal Pendidikan Sains Indonesia, 10*(2), 342–354. https://doi.org/10.24815/jpsi.v10i2.23844

Martawijaya, M. A., Rahmadhanningsih, S., Swandi, A., Hasyim, M., & Sujiono, E. H. (2023). The effect of applying the ethno-STEM-project-based learning model on students' higher-order thinking skill and misconception of physics topics related to Lake Tempe, Indonesia. *Jurnal Pendidikan IPA Indonesia, 12*(1), 1–13. https://doi.org/10.15294/jpii.v12i1.38703

Mudi, Y. J., Hidayati, K., Nursa'ban, M., & Pusporini, W. (2026). Mapping the landscape of critical thinking assessment in STEM education: A systematic review of psychometric properties, contextual implementation, and future directions. *Integrated Science Education Journal, 7*(2), 265–274. https://doi.org/10.37251/isej.v7i2.2385

Nasriya, T. Z., Wuryandani, W., & Mulyono. (2021). Explosion box: A learning media of ecosystem components for elementary school student. *Jurnal Pendidikan Progresif, 11*(2), 338–348. https://doi.org/10.23960/jpp.v11.i2.202115

Noperi, H., Diani, R., Irwandani, I., Suryani, Y., Sodikin, S., & Saputro, T. (2025). Psychometric analysis of the critical thinking skills instrument in physics learning using the Rasch model. *Indonesian Journal of Science and Mathematics Education, 8*(3), 627–641. https://doi.org/10.24042/ijsme.v8i3.28770

Nurhayani, N., Lubis, R., & Yusnaldi, E. (2024). Pengaruh media explosion box terhadap kemampuan berpikir kritis siswa pada mata pelajaran IPS kelas V MIN 5 Labuhanbatu. *Jurnal Sains Student Research, 2*(3), 984–995. https://doi.org/10.61722/jssr.v2i3.1861

Organisation for Economic Co-operation and Development. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Perry, K., Meissel, K., & Hill, M. F. (2022). Rebooting assessment: Exploring the challenges and benefits of shifting from pen-and-paper to computer in summative assessment. *Educational Research Review, 36*, Article 100451. https://doi.org/10.1016/j.edurev.2022.100451

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what is being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497. https://doi.org/10.1002/nur.20147

Primasari, R., Miarsyah, M., & Rusdi, R. (2020). Science literacy, critical thinking skill, and motivation: A correlational study. *JPBI (Jurnal Pendidikan Biologi Indonesia), 6*(2), 273–282. https://doi.org/10.22219/jpbi.v6i2.11124

Putri, R. M., & Asrizal. (2023). Need analysis of developing digital teaching materials to improve 21st century skills. *Jurnal IPA dan Pembelajaran IPA, 7*(2), 108–117. https://doi.org/10.24815/jipi.v7i2.29797

Ramdani, A., Jufri, A. W., Gunawan, G., Fahrurrozi, M., & Yustiqvar, M. (2021). Analysis of students' critical thinking skills in terms of gender using science teaching materials based on the 5E learning cycle integrated with local wisdom. *Jurnal Pendidikan IPA Indonesia, 10*(2), 187–199. https://doi.org/10.15294/jpii.v10i2.29956

Rodríguez-Rojas, M.-P., Ramírez-Segura, A.-M., Valenzuela-Mora, P., Dampuré, J., & Riveros Munévar, F. (2024). Critical thinking evaluation scale: Design and validation in a Colombian population. *SAGE Open, 14*(4), Article 21582440241297418. https://doi.org/10.1177/21582440241297418

Romero Jeldres, M., Díaz Costa, E., & Faouzi Nadim, T. (2023). A review of Lawshe's method for calculating content validity in the social sciences. *Frontiers in Education, 8*, Article 1271335. https://doi.org/10.3389/feduc.2023.1271335

Safitri, A. R. R., Putra, A. P., & Mudiono, A. (2021, June). Development of 3-dimensional thematic explosion box (TEP) media with strengthening the character of curiosity in class II of elementary school. In *Proceedings of the 4th International Conference on Education 2021* (p. 140).

Saimon, M., Lavicza, Z., & Dana-Picard, T. (2023). Enhancing the 4Cs among college students of a communication skills course in Tanzania through a project-based learning model. *Education and Information Technologies, 28*(6), 6269–6285. https://doi.org/10.1007/s10639-022-11406-9

Samura, A. O., & Darhim. (2023). Improving mathematics critical thinking skills of junior high school students using blended learning model in GeoGebra-assisted mathematics learning. *International Journal of Interactive Mobile Technologies, 17*(2), 101–117. https://doi.org/10.3991/ijim.v17i02.36097

Saputra, B., Dharma, A. K., & Kusuma, G. P. (2024). Keanekaragaman hayati sebagai sumber belajar kontekstual dalam kurikulum pendidikan biologi. *Jurnal Cakrawala Pendidikan dan Biologi, 1*(3), 7–17. https://doi.org/10.61132/jucapenbi.v1i3.181

Stefana, A., Damiani, S., Granziol, U., Provenzani, U., Solmi, M., Youngstrom, E. A., & Fusar-Poli, P. (2025). Psychological, psychiatric, and behavioral sciences measurement scales: Best practice guidelines for their development and validation. *Frontiers in Psychology, 15*, Article 1494261. https://doi.org/10.3389/fpsyg.2024.1494261

Sugiyono. (2013). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

Sujanem, R., & Suwindra, I. P. (2023). Problem-based interactive physics e-module in physics learning through blended PBL to enhance students' critical thinking skills. *Jurnal Pendidikan IPA Indonesia, 12*(1), 135–145. https://doi.org/10.15294/jpii.v12i1.39971

Sumarni, W., & Kadarwati, S. (2020). Ethno-STEM project-based learning: Its impact to critical and creative thinking skills. *Jurnal Pendidikan IPA Indonesia, 9*(1), 11–21. https://doi.org/10.15294/jpii.v9i1.21754

Suprapto, E., Saryanto, S., Sumiharsono, R., & Ramadhan, S. (2020). The analysis of instrument quality to measure the students' higher order thinking skill in physics learning. *Journal of Turkish Science Education, 17*(4), 520–527. https://doi.org/10.36681/tused.2020.42

Suryawati, E., Suzanti, F., Zulfarina, Z., Putriana, A. R., & Febrianti, L. (2020). The implementation of local environmental problem-based learning student worksheets to strengthen environmental literacy. *Jurnal Pendidikan IPA Indonesia, 9*(2), 169–178. https://doi.org/10.15294/jpii.v9i2.22892

Tafakur, T., Retnawati, H., & Shukri, A. A. M. (2023). Effectiveness of project-based learning for enhancing students' critical thinking skills: A meta-analysis. *JINoP (Jurnal Inovasi Pembelajaran), 9*(2), 191–209. https://doi.org/10.22219/jinop.v9i2.22142

Tanti, T., Kurniawan, D. A., Kuswanto, K., Utami, W., & Wardhana, I. (2020). Science process skills and critical thinking in science: Urban and rural disparity. *Jurnal Pendidikan IPA Indonesia, 9*(4), 489–498. https://doi.org/10.15294/jpii.v9i4.24139

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Leadership Training Institute/Special Education, University of Minnesota. https://files.eric.ed.gov/fulltext/ED090725.pdf

van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. (2020). Determinants of 21st-century skills and 21st-century digital skills for workers: A systematic literature review. *SAGE Open, 10*(1), Article 2158244019900176. https://doi.org/10.1177/2158244019900176

Verawati, N. N. S. P., Hikmawati, H., & Prayogi, S. (2020). The effectiveness of inquiry learning models intervened by reflective processes to promote critical thinking ability in terms of cognitive style. *International Journal of Emerging Technologies in Learning, 15*(16), 212–220. https://doi.org/10.3991/ijet.v15i16.14687

Verawati, N. N. S. P., Hikmawati, H., Prayogi, S., & Bilad, M. R. (2021). Reflective practices in inquiry learning: Its effectiveness in training pre-service teachers' critical thinking viewed from cognitive styles. *Jurnal Pendidikan IPA Indonesia, 10*(4), 505–514. https://doi.org/10.15294/jpii.v10i4.31814

Wayudi, W., Supriadi, S., & Ramdani, A. (2020). Pengaruh model pembelajaran berbasis masalah terhadap kemampuan berpikir kritis siswa SMA Negeri Bandung. *Jurnal Pendidikan Sains Indonesia, 8*(2), 144–155. https://doi.org/10.24815/jpsi.v8i2.17632

Wulandari, D. T., Diani, R., & Noperi, H. (2025). Development of diagnostic instruments for critical thinking skills of students in physics learning on socio-scientific issues. *Journal Corner of Education, Linguistics, and Literature, 5*(1), 94–112. https://doi.org/10.54012/jcell.v5i1.534

Yanuarto, W. N., Jaelani, A., & Purwanto, J. (2021). Flipped classroom model: Empowering digital literacy for mathematics learning in Society 5.0. *Indonesian Journal of Science and Mathematics Education, 4*(2), 158–171. https://doi.org/10.24042/ijsme.v4i2.9638

Yunita, Y., Diani, R., Noperi, H., & Velina, Y. (2024). Development of interactive e-modules with STEM approach using Canva application to improve students' critical thinking ability in physics learning. *JINoP (Jurnal Inovasi Pembelajaran), 10*(2), 192–210. https://doi.org/10.22219/jinop.v10i2.32004

Zen, Z., Reflianto, R., Syamsuar, S., & Ariani, F. (2022). Academic achievement: The effect of project-based online learning method and student engagement. *Heliyon, 8*(11), Article e11509. https://doi.org/10.1016/j.heliyon.2022.e11509

Zulaiha, Z., Hartono, H., & Siahaan, S. M. (2025). Needs analysis for the development of web-based learning multimedia context climate change issues environmental issues content to support junior high school students' literacy. *Journal of Innovation and Research in Primary Education, 4*(3), 1481–1490. https://doi.org/10.56916/jirpe.v4i3.1596