

Critical Thinking Skills in Elementary School Science Lessons

Anugrah Dwicahyandi Oktavian ¹, Rusijono², Fajar Arianto³

^{1,2,3} Master's Program in Educational Technology, Universitas Negeri Surabaya, Indonesia.

 24010905012@mhs.unesa.ac.id

Abstract. This study aims to design and test the validity and reliability of a critical thinking assessment instrument based on Ennis's theory, applied to the IPAS subject theme Exploring the World. The instrument was developed in the form of essay questions covering six critical thinking indicators: clarification, assumption, inference, evaluation, explanation, and self-regulation. The research adopted a Research and Development (R&D) approach using the Borg & Gall model. The try-out subjects were 30 sixth-grade students from Elementary School Pucangro Jombang. Validity was evaluated by three education and IPAS learning experts using Aiken's V, while reliability was tested using Cronbach's Alpha. Results showed that all items had high validity (Aiken's V > 0.82) and strong reliability ($\alpha = 0.824$). This instrument is considered appropriate as an assessment tool for evaluating elementary students' critical thinking skills in a thematic and global learning context. The use of the Exploring the World theme proved effective in encouraging students to analyze, compare, and draw critical conclusions about cultures, environments, and global life.

Keywords: Critical Thinking; Ennis; IPAS; Exploring World; Validity; Reliability.

1. INTRODUCTION

One of the key competencies in 21st-century education that is the focus of numerous national and international education initiatives is critical thinking. Critical thinking is positioned as a fundamental skill that needs to be cultivated starting in primary school in Indonesia's Merdeka Curriculum in order to produce a generation that can think critically, logically, and empirically when confronted with the challenges of a global society (Kemendikbudristek, 2022). The teaching of Natural and Social Sciences (IPAS), which places a strong emphasis on the interdisciplinary integration of science, social studies, and environmental studies, makes this ability extremely pertinent. IPAS learning resources, such the "Traveling Around the World" theme, give students plenty of opportunity to investigate global issues from a critical standpoint, including contrasting national cultures, climates, and technological advancements (Qazi & Al-Mhdawi, 2024; Firoozi et al., 2025). Therefore, adequate, relevant, and standardized assessment tools are needed to integrate critical thinking skills into IPAS.

Although there have been numerous calls to develop critical thinking skills, implementing this practice at the elementary school level continues to face several challenges. The assessment instruments used in the classroom are typically based on factual and illustrative knowledge and are not well-suited to exploring students' higher-order thinking processes (Radović et al., 2024). However, effective

assessments do more than just reducing learning outcomes; they are also educational tools that encourage students to reflect and evaluate their own learning process (Chang & Sun, 2024). The lack of assessment instruments specifically developed based on established critical thinking frameworks, such as Ennis's model, represents a fundamental issue that must be addressed through systematic research aligned with the needs of school-based education (Makaruk et al., 2024).

Various current studies show that using an assessment approach based on critical theory can significantly improve the quality of education. However, most of the existing instrument development remains focused on secondary or higher education and does not fully capture the context of thematic and exploratory learning in elementary schools. This condition creates a gap in literature where there is a significant need for the design of critical thinking tools that are relevant in a contextual setting, can be used in IPAS education, and have reliability and validity that can be tested (Al-kfairy et al., 2024).

The innovation of this study lies in the development of a critical thinking assessment instrument based on the theory of Sanabria-Z et al. (2024), specifically contextualized within the "Pelesir Keliling Dunia" lesson in the Grade VI IPAS curriculum at the elementary school level. This study not only identifies critical indicators such as clarity, inference, evaluation, and self-awareness, but it also highlights the importance of global awareness and sensitivity that are discussed in educational topics (Salehian et al., 2023). In addition, the use of clearly defined topics and an indicator-based rubric enhances analytical clarity and encourages students to develop arguments and engage in reflective writing more thoroughly. As a remedy for existing problems, research and development (R&D) is used in this study to produce valid, systematic, and reliable instruments for analysis in the course of execution. Validity will be assessed using the Aiken's V test, while reliability will be assessed using Cronbach's Alpha once a more thorough test is administered to the students (Rothschild et al., 2009). Accordingly, the final product developed is not only theoretically valuable for analyzing the literature on elementary education assessment, but also offers practical contributions to teachers in assessing higher-order thinking skills in a reliable and consistent manner (Wu et al., 2024).

Based on the background described above, this study aims to evaluate and develop an assessment instrument for measuring elementary school students' critical thinking skills, grounded in Ennis's theory and contextualized within the IPAS lessons on the 'Travel Around the World' theme. It is hoped that this instrument can be used as a *sahih andal* tool to help teachers identify, assess, and improve students' critical thinking skills while also enhancing the implementation of theme-based learning that is focused on 21st-century skills (Carretta et al., 2014). Furthermore, the findings of this study are expected to contribute to the refinement of assessment practices in primary education, particularly in fostering

reflective, analytical, and globally oriented learners.

2. METHOD

This research employs the Research and Development (R&D) paradigm by focusing on the Borg and Gall development model. This model is designed to develop educational products such as curricula, learning materials, or assessment tools through systematic steps that link research and product development. In the context of this study, the product being developed is an assessment instrument for evaluating critical thinking skills among elementary school students. It is specifically designed based on Ennis theory and contextualized with the IPAS curriculum for “Keliling Dunia”. R&D is an effective method since it not only produces a product but also ensures that it is valid in theory and practice through ahli and lapangan tests (Borg & Gall, 1983; Sugiyono, 2016).

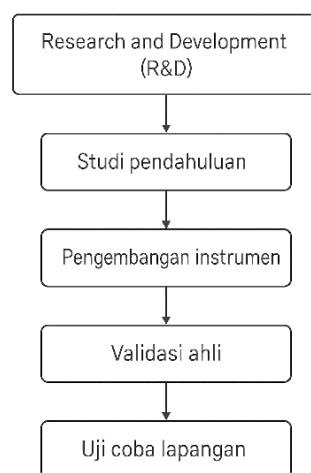


Figure 1. Research Flow Chart

This method's goal is to ensure that the instruments being used are not only accurate in assessing critical thinking skills but also relevant to the learning context and easy for teachers to use. Therefore, the Research and Development (R&D) approach offers flexibility in monitoring, evaluating, and refining instruments based on feedback from students and field users, resulting in products that are effective, efficient, and applicable within the context of elementary education (Borg & Gall, 1983; Plomp & Nieveen, 2013).

The study's subjects are 30 sixth-grade students from Elementary School Pucangro Jombang who were chosen on purpose since they had already received materials from “Pelesir Keliling Dunia” in IPAS classes. In addition to this, three ahli are involved in the instrument validation process: An expert in educational evaluation, a sixth-grade teacher with classroom experience, and a specialist in elementary-level IPAS. The purpose of involving these experts is to ensure that the instruments used are not only conceptually sound but also pedagogically appropriate within the context of classroom teaching (Nitko &

Brookhart, 2014).

The instruments developed took the form of open-ended questions and assessment rubrics based on Ennis' six indicators of critical thinking, namely: (1) clarification, for example, "Why do countries in the polar regions have smaller populations?"; (2) assumptions, such as "What assumptions do we make when we think that all developed countries must have modern transportation systems?"; (3) inference, such as "Based on temperature and rainfall data, which country is suitable for agriculture?"; (4) evaluation, for example, "Compare the lifestyles of Japanese and Indonesian societies. What are the positive and negative aspects?"; (5) explanation, such as "Explain why a country has become a global tourist destination"; and (6) self-regulation, for example, "Revise your opinion after reading new data about the African continent." Each question is designed to assess critical thinking skills in depth, aligned with the IPAS theme context and 21st-century skills (Facione, 2015; Kemendikbudristek, 2022).

The content validity of the instrument was tested using Aiken's V formula, involving the three experts as assessors. An Aiken's V value of more than 0.8 is considered to indicate high validity (Azwar, 2015). Meanwhile, reliability was tested using Cronbach's Alpha coefficient from the results of trials involving 30 students. The instrument was deemed reliable if the α value was greater than 0.7, indicating good internal consistency between items in measuring the same ability (Arikunto, 2019). The results of the validity and reliability analysis form the basis for determining the suitability of the instrument for widespread use in assessing critical thinking in elementary school students.

3. RESULTS

This research begins with a study of literature that examines critical thinking theories, IPAS curricula, and the need for assessment tools that are appropriate for the highest levels of schooling. This study shows that even though critical thinking is a skill that is highly valued in abad ke-21 education, the available asesmen instruments are mostly focused on factual and heuristic knowledge rather than the process of critical thinking that is necessary to develop students' analytical skills. Because of this, it is very necessary to develop instruments that are more accurate and reliable for assessing students' critical thinking skills at the highest level of dasar school.

The development of a preliminary draft of the instrument is the next step. The instruments used in this study are based on the six indicators of critical thinking according to Ennis, which include clarification, assumption, inference, evaluation, clarification, and self-regulation. These topics are intended to help students learn more about a variety of global issues related to IPAS education, such as the differences in the world's climate, the relationship between geography and the general population's way of life, and the effects of social and technological change. These soal-soal are also intended to assess students' abilities in evaluating

information, expressing their thoughts clearly, and explaining their views or reasons.

The next step is for the ahli to validate the instrument. This process involves three competent educational experts: a specialist in educational evaluation, an experienced sixth-grade teacher, and a subject-matter expert in elementary-level IPAS instruction. The purpose of this article is to determine whether the instrument is in line with the theoretical indicators that are critical and pertinent to the IPAS learning context that addresses global issues. According to the validation results, this instrument meets the criteria for high validity, with an Aiken's V value of 0.82 for all test items. This indicates that the instrument aligns with theoretical indicators and is suitable for use in real classroom settings in elementary schools.

Following validation, 30 students from grade VI Elementary School Pucangro Jombang were given the instrument. The purpose of this trial is to assess the effectiveness of the instrument in evaluating students' critical thinking skills. Based on the aforementioned trial, the instrument is easy for students to understand and can be effectively used to measure critical thinking skills. Students are able to provide insightful responses to the essay topics being discussed, and this instrument is capable of generating data that clearly reflects their critical thinking abilities.

Reliability testing is also done to ensure internal instrument consistency. The reliability analysis using Cronbach's Alpha yielded a value of 0.824, indicating that the instrument has very good internal consistency. This indicates that the instrument can be used to assess students' critical thinking skills without bias or inaccuracy in measurement.

Table 1. Content Validity Results of the Instrument

Question	Aiken's V	Description
Question 1	0.85	Valid High
Question 2	0.83	Valid High
Question 3	0.84	Valid High
Question 4	0.82	Valid High
Question 5	0.87	Valid High
Question 6	0.86	Valid High

Table 2. Instrument Reliability Test Results

Indicators	Cronbach's Alpha Value	Description
Clarification	0.80	Consistent and Valid
Assumptions	0.82	Consistent and Valid
Inferences	0.81	Consistent and Valid
Evaluation	0.85	Consistent and Valid
Explanation	0.83	Consistent and Valid
Self-regulation	0.84	Consistent and Valid

The overall results of the school-based trial indicate that the instruments used demonstrate very good validity and reliability. This instrument successfully improves students' critical thinking skills and can be used practically by teachers in evaluating critical thinking in IPAS classes. The lessons incorporated in this instrument can help students become more critical and reflective about various global issues, while also supporting their understanding of the higher-order thinking skills expected in the 21st-century curriculum. This instrument is suitable for use in assessing students' critical thinking skills in a dasar school with a theme-based and interdisciplinary learning environment.

4. DISCUSSION

The results of this study indicate that the instruments used to assess students' critical thinking skills in dasar schools have been proven to be reliable and valid. This instrument, which focuses on the six indicators of critical thinking, according to Ennis's framework, is designed to guide students in developing analytical, evaluative, and reflective thinking skills to Nemati-Vakilabad et al., (2023), can provide clear illustrations of students' critical thinking abilities in the context of IPAS education. This study demonstrates that students may easily understand the topics being covered and provide feedback that strengthens their analytical and reflective skills. The results are consistent with previous research that indicates that instruments based on critical thinking indicators can increase students' ability to think critically and analytically (Zhang et al., 2025). However, there is a difference with research by Leal Filho et al., (2024), which states that students at the top of the school are frequently unable in applying critical thinking in more complex contexts.

The results of this validation show that the instrument has met the very good validity criteria, with an Aiken's V value of 0.82 for all subject matter. This nilai highlights the high degree of agreement between the theoretical indicator based on critical analysis and the topic that is developed in the instrument. This supports the findings of previous research that indicates that high-quality test validity is crucial for ensuring that the asesmen instrument detects what should be checked, namely critical thinking skills (LeMire et al., 2024). However, according to a few previous studies, credible instruments consistently reduce difficulties in their application in the field, especially when used with students from areas where literacy levels are lower (Özer & Şahin Altun, 2024). According to this study, students in grade VI Elementary School can understand and respond appropriately to the aforementioned topics.

Regarding reliability results, which show a Cronbach's Alpha value of 0.824, this study confirms that the instrument being used has high internal consistency. This result is also consistent with research by Buin et al., (2023), which states that instruments with Cronbach's Alpha values more than 0.7 can be considered reliable for use in educational research. On the other hand, some recent studies,

such as those conducted by Vipu Vinayak et al., (2024), indicate that tools designed to assess critical thinking skills in dasar schools have lower reliability due to difficulties in designing lessons that match students' cognitive levels. This study shows that by analyzing the terbuka soal-soal uraian according to the Ennis indicator, the instrument may produce reliable and consistent data.

Although this instrument has been proven to be valid and reliable, it was found that there are several challenges that need to be overcome, especially in terms of cultural context. Instrument validation conducted by experts showed recommendations to avoid cultural stereotypes in questions comparing the lives of people between countries. This highlights that cultural sensitivity is an important factor that must be considered in the design of critical thinking assessment instruments, especially when such instruments are used in multicultural contexts (Göğebakan Yıldız & Atman Uslu, 2024). Although this instrument has already accommodated these needs, the challenge of creating questions that are fully neutral and culturally sensitive remains a key consideration in the development of future assessment instruments.

The implications of this study are very important for primary education in Indonesia, especially in developing more comprehensive assessment instruments based on critical thinking skills. This study proves that Ennis' critical thinking theory-based instruments can be effectively applied in primary schools, and the results can provide a more in-depth picture of students' higher-order thinking skills. The use of this instrument can help teachers design learning that focuses more on the thinking process, not just on learning outcomes. Therefore, this instrument can be used not only as an assessment tool but also as a means to improve the quality of 21st-century competency-based learning.

5. CONCLUSION

The critical thinking assessment instrument developed using Ennis' theory and contextualized within the theme of World Travel has proven to have excellent validity and reliability for use at the elementary school level. This instrument successfully measures students' critical thinking skills in a deep and reflective manner, leading to the development of higher-order thinking skills in line with the objectives of the Merdeka Curriculum. Additionally, this instrument is reliable in providing an accurate picture of students' critical thinking skill development, with Aiken's V values indicating high validity and Cronbach's Alpha values confirming good internal consistency. However, challenges related to cultural sensitivity in some questions need to be addressed in further instrument development. Therefore, it is recommended that this instrument be widely implemented in elementary schools as an assessment tool that not only measures factual knowledge but also more complex critical thinking skills, while considering cultural context factors in question development. Furthermore, it is important to integrate this instrument into formative assessment activities, so that

it can help teachers monitor students' critical thinking skills development continuously and provide constructive feedback for more effective learning.

6. REFERENCES

- Al-kfairy, M., Ahmed, S., & Khalil, A. (2024). Factors impacting users' willingness to adopt and utilize the metaverse in education: A systematic review. *Computers in Human Behavior Reports*, 15, 100459. <https://doi.org/10.1016/j.chbr.2024.100459>
- Arikunto, S. (2019). *Prosedur penelitian: Suatu pendekatan praktik (Edisi Revisi)*. Rineka Cipta.
- Azwar, S. (2015). *Reliabilitas dan validitas*. Pustaka Pelajar.
- Borg, W. R., & Gall, M. D. (1983). *Educational research: An introduction (4th ed.)*. Longman.
- Buin, E., da Mota Silveira, M. S. V., & Pavin, E. J. (2023). Adaptation and psychometric assessment of the instrument "Partners of adults with type 1 diabetes distress scale" in a Brazilian population. *PEC Innovation*, 2, 100155. <https://doi.org/10.1016/j.pecinn.2023.100155>
- Carretta, C. M., Ridner, S. H., & Dietrich, M. S. (2014). Hope, hopelessness, and anxiety: A pilot instrument comparison study. *Archives of Psychiatric Nursing*, 28(4), 230–234. <https://doi.org/10.1016/j.apnu.2014.05.005>
- Chang, W.-L., & Sun, J. C.-Y. (2024). Evaluating AI's impact on self-regulated language learning: A systematic review. *System*, 126, 103484. <https://doi.org/10.1016/j.system.2024.103484>
- Facione, P. (2015). *Critical thinking: What it is and why it counts*. Insight Assessment.
- Firoozi, A. A., Firoozi, A. A., & Maghami, M. R. (2025). Harnessing photovoltaic innovation: Advancements, challenges, and strategic pathways for sustainable global development. *Energy Conversion and Management: X*, 27, 101058. <https://doi.org/10.1016/j.ecmx.2025.101058>
- Göğebakan Yıldız, D., & Atman Uslu, N. (2024). Investigating the effect of digital storytelling on multicultural competencies and social justice: A mixed method study in psychological counseling education. *International Journal of Intercultural Relations*, 103, 102065. <https://doi.org/10.1016/j.ijintrel.2024.102065>
- Kemendikbudristek. (2022). *Buku guru IPAS kelas VI SD kurikulum merdeka*. Pusat Kurikulum dan Perbukuan.
- Leal Filho, W., Schmidberger, I., Sharifi, A., Vargas, V. R., Rampasso, I. S., Dibbern, T., Liakh, O., Aina, Y. A., Trevisan, L. V., Mbah, M. F., Anholon, R., & Kozlova, V. (2024). Design thinking for sustainable development: A bibliometric analysis and case study research. *Journal of Cleaner Production*, 455, 142285. <https://doi.org/10.1016/j.jclepro.2024.142285>
- LeMire, S., Xu, Z., Dorsey, L. G., & Hahn, D. (2024). Information literacy skill mastery across the undergraduate degree: An examination of first-generation and

- continuing-generation students. *The Journal of Academic Librarianship*, 50(5), 102932. <https://doi.org/10.1016/j.acalib.2024.102932>
- Makaruk, H., Webster, E. K., Porter, J., Makaruk, B., Bodasińska, A., Zieliński, J., Tomaszewski, P., Nogal, M., Szyszka, P., Starzak, M., Śliwa, M., Banaś, M., Biegajło, M., Chaliburda, A., Gierczuk, D., Suchecki, B., Molik, B., & Sadowski, J. (2024). The fundamental motor skill proficiency among Polish primary school-aged children: A nationally representative surveillance study. *Journal of Science and Medicine in Sport*, 27(4), 243–249. <https://doi.org/10.1016/j.jsams.2023.12.007>
- Nemati-Vakilabad, R., Mojebi, M. R., Mostafazadeh, P., Jafari, M. J., Kamblash, A. J., Shafaghat, A., Abbasi, A. S., & Mirzaei, A. (2023). Factors associated with the critical thinking ability among nursing students: An exploratory study in Iran. *Nurse Education in Practice*, 73, 103814. <https://doi.org/10.1016/j.nepr.2023.103814>
- Nitko, A. J., & Brookhart, S. M. (2014). *Educational assessment of students* (7th ed.). Pearson.
- Özer, D., & Şahin Altun, Ö. (2024). Nursing students' mental health literacy and resilience levels: A cross-sectional study. *Archives of Psychiatric Nursing*, 51, 222–227. <https://doi.org/10.1016/j.apnu.2024.06.017>
- Plomp, T., & Nieveen, N. (2013). *Educational design research*. SLO.
- Qazi, A., & Al-Mhdawi, M. K. S. (2024). Exploring critical drivers of global innovation: A bayesian network perspective. *Knowledge-Based Systems*, 299, 112127. <https://doi.org/10.1016/j.knosys.2024.112127>
- Radović, S., Seidel, N., Menze, D., & Kasakowskij, R. (2024). Investigating the effects of different levels of students' regulation support on learning process and outcome: In search of the optimal level of support for self-regulated learning. *Computers & Education*, 215, 105041. <https://doi.org/10.1016/j.compedu.2024.105041>
- Rothschild, Z. K., Abdollahi, A., & Pyszczynski, T. (2009). Does peace have a prayer? The effect of mortality salience, compassionate values, and religious fundamentalism on hostility toward out-groups. *Journal of Experimental Social Psychology*, 45(4), 816–827. <https://doi.org/10.1016/j.jesp.2009.05.016>
- Salehian, A., Aghabayk, K., Seyfi, M., & Shiwakoti, N. (2023). Comparative analysis of pedestrian crash severity at United Kingdom rural road intersections and Non-Intersections using latent class clustering and ordered probit model. *Accident Analysis & Prevention*, 192, 107231. <https://doi.org/10.1016/j.aap.2023.107231>
- Sanabria-Z, J., Cruz-Sandoval, M., Moreno-Romo, A., Bosch-Gómez, S., & Ramírez-Montoya, M. S. (2024). Research foresight in bridging open science and open innovation: Overview based on the complex thinking paradigm. *International Journal of Innovation Studies*, 8(1), 59–75. <https://doi.org/10.1016/j.ijis.2023.08.002>

- Sugiyono. (2016). Metode penelitian kuantitatif, kualitatif dan R&D. Alfabeta.
- Vipu Vinayak, V. J., Deshmukh, K., Murthy, V. R. K., & Pasha, S. K. K. (2024). Conducting polymer based nanocomposites for supercapacitor applications: A review of recent advances, challenges and future prospects. *Journal of Energy Storage*, 100, 113551. <https://doi.org/10.1016/j.est.2024.113551>
- Wu, T.-T., Sari, N. A. R. M., & Huang, Y.-M. (2024). Integrating extended formative assessment in flipped jigsaw learning: Promoting learning engagement and higher-order thinking skills in international business education context. *The International Journal of Management Education*, 22(1), 100930. <https://doi.org/10.1016/j.ijme.2024.100930>
- Zhang, Y., Guo, W., Ji, B., & Shi, H. (2025). An exploratory case study of Chinese college students' critical thinking skills through interactive independent courses: evidence from behavioral and brain imaging. *Thinking Skills and Creativity*, 57, 101867. <https://doi.org/10.1016/j.tsc.2025.101867>