JPPIPA 11(8) (2025)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

Development and Implementation of a Custom GPT Model for Assessing Critical Thinking Skills in Higher Education

Muhammad Fathurrahman[1*], Faisal Eka Mahendra[1], Karim Paul Aronggear[1], Muhammad Syahrul Kahar[1], Teguh Hidayat Iskandar Alam[2]

[1] Mathematics Education Department, Universitas Muhammadiyah Sorong, Sorong, Indonesia.
[2] Information Technology Department, Universitas Muhammadiyah Sorong, Sorong, Indonesia.

Received: May 16, 2025
Revised: July 08, 2025
Accepted: August 25, 2025
Published: August 31, 2025

Corresponding Author:
Muhammad Fathurrahman
r.fathur.ums@gmail.com

DOI: 10.29303/jppipa.v11i8.12067

Abstract: This study developed and evaluated a Custom GPT model designed specifically to assess the critical-thinking skills of higher-education students, focusing on synthesis (C4), evaluation (C5), and creation (C6) based on Bloom's Taxonomy. Employing a design-based research (DBR) methodology, six iterative design cycles were systematically conducted, involving direct classroom observations, expert validations, and practical testing with fourteen mathematics education students. Findings demonstrated that the Custom GPT system is highly practical regarding ease of use (86.7%) and efficiency (85.1%), and practically engaging for learners (81.8%). Quantitative analyses showed significant improvements in students' critical-thinking skills, particularly within synthesis and evaluation domains, although variations in creation domain performance indicated the need for incorporating more real-world, application-based scenarios. Qualitative feedback confirmed that real-time AI-driven feedback significantly enhanced students' self-reflection and learning strategies. This study contributes theoretically by integrating AI with Bloom's Taxonomy-based assessment and offers practical implications by providing a scalable, objective, and innovative assessment approach for enhancing higher-order thinking in higher education contexts.

Keywords: Artificial Intelligence; Bloom's taxonomy; Critical thinking; Custom GPT.

Introduction

Rapid advances in educational technology have presented both challenges and opportunities for enhancing students' critical-thinking skills (Anisah et al., 2025). Critical thinking, defined as the ability to analyse, evaluate, and synthesise information, is essential for problem-solving and decision-making in today's dynamic and complex global environment (Elen & Verburgh, 2023; Kadrija et al., 2022). Despite its importance, fostering and assessing critical thinking remains a significant challenge, particularly in higher education, where the development of such skills is often inconsistent across disciplines. The 2022 Programme for International Student Assessment (PISA) revealed a mixed picture of educational progress in Indonesia (Cahyani & Setiawan, 2024). While the country's rankings in literacy, mathematics, and science improved, its average scores remained below the Organisation for Economic Co-operation and Development (OECD) standard, highlighting the need for innovative strategies to enhance higher-order thinking skills (OECD, 2024).

The integration of artificial-intelligence (AI) technology, particularly models such as the Generative Pre-trained Transformer (GPT) offers a promising

How to Cite:
Fathurrahman, M., Mahendra, F. E., Aronggear, K. P., Kahar, M. S., & Alam, T. H. I. (2025). Development and Implementation of a Custom GPT Model for Assessing Critical Thinking Skills in Higher Education. *Jurnal Penelitian Pendidikan IPA*, *11*(8), 169–176. https://doi.org/10.29303/jppipa.v11i8.12067

avenue for addressing these challenges. AI's capacity to process large datasets and deliver real-time feedback aligns with the cognitive demands of critical-thinking education (Bitzenbauer, 2023; van den Berg & du Plessis, 2023). By leveraging AI for personalised learning and assessment, educators can create more inclusive and effective learning environments, enabling students to engage with content that fosters higher-order thinking. This approach not only addresses the gaps identified in educational assessment but also prepares students for the rapidly evolving workforce, in which critical thinking is a core competency (Indrašienė et al., 2023; López et al., 2023).

Recent research highlights the significant potential of integrating artificial intelligence (AI) in fostering and assessing critical-thinking skills. For example, Bitzenbauer (2023) found that AI-based models such as ChatGPT can be employed to design learning activities that promote critical thinking through adaptive and interactive approaches. Furthermore, van den Berg & du Plessis (2023) emphasize that AI applications can deliver personalized feedback, enabling students to identify their strengths and weaknesses in real time. These findings align with López et al. (2023), who demonstrated that applying AI in vocational education significantly enhances students' analytical and evaluative skills through project-based learning environments.

Approaches that combine AI with traditional pedagogical strategies have also proven effective. For instance, Surin & Damrongpanit (2024) reported that integrating AI into meta-analysis–based learning management reinforces students' critical-thinking skills by providing a structured framework for exploration and reflection. Meanwhile, Hopfenbeck et al. (2023) acknowledged the challenges of applying AI to formative assessment but likewise highlighted its potential to yield deep insights into students' thinking processes (Andiopenta, 2024). Collectively, these studies offer a robust theoretical foundation for leveraging AI, particularly models such as GPT to assess critical-thinking abilities comprehensively and effectively.

This study introduces an innovative approach to developing and implementing a Generative Pre-trained Transformer (GPT) model specifically designed to holistically evaluate university students' critical-thinking skills. Unlike previous research, which has largely focused on AI as a learning aid (López et al., 2023; van den Berg & du Plessis, 2023), the present work contributes by presenting an AI-based evaluation system capable of assessing the synthesis, evaluation, and creation facets in line with Bloom's taxonomy. The system not only delivers quantitative assessments but also provides qualitative insights into students' critical-thinking patterns, making it highly relevant for curriculum development in higher education.

Moreover, this study broadens the theoretical framework by integrating a real-time, AI-based assessment method that delivers immediate feedback an approach seldom applied in previous literature (Hopfenbeck et al., 2023; Surin & Damrongpanit, 2024). By combining AI's analytical sophistication with educational-assessment design, the research bridges gaps in the literature and offers a practical model adaptable to diverse educational contexts. This contribution is highly significant for institutions seeking to enhance learning quality through modern, data-driven technological approaches.

This study aims to develop and evaluate an artificial-intelligence-based assessment model specifically, a customized Generative Pre-trained Transformer (Custom GPT), to measure higher-education students' critical-thinking skills, with a primary focus on synthesis (C4), evaluation (C5), and creation (C6) according to Bloom's Taxonomy. Recent research indicates that conventional methods for assessing critical-thinking skills are increasingly inefficient in evaluating complex learning achievements, underscoring the need for more innovative and effective AI-based evaluation systems (Xiaolei & Teng, 2024).

Specifically, this study has several primary objectives. First, it seeks to identify the need for an AI-based system to evaluate critical-thinking skills among students in the Mathematics Education program through an in-depth needs analysis, including classroom observations, interviews with lecturers and students, and analysis of learning documents. This effort aims to address the challenges of critical-thinking assessment, which is currently insufficiently objective and difficult to scale within higher-education environments (Xia et al., 2024).

Second, this study aims to develop a critical-thinking assessment instrument in the form of a Custom GPT model capable of objective evaluation and real-time feedback. The use of GPT in higher-education contexts has proven effective in enhancing students' higher-order thinking abilities, such as analysis, evaluation, and creation (Kofahi et al., 2025; Lee et al., 2024).

Third, through a Design-Based Research (DBR) approach, this study carries out iterative testing and refinement of the Custom GPT model in real classroom settings. The DBR approach has proven effective in producing technology-based educational interventions that are contextual, practical, and better able to address educational challenges (Hoadley & Campos, 2022).

Fourth, this study will analyse the impact of using the Custom GPT model on improving the quality of the learning process, particularly in supporting the

development of students' critical-thinking skills. The integration of AI such as GPT into learning has been proven to provide significant real-time feedback, enhance students' self-reflection, and, overall, improve learning outcomes and student engagement in the educational process (Lee et al., 2024; Swindell et al., 2024).

Thus, this study is expected to yield an innovative and academically relevant evaluation model that makes a tangible contribution to the higher-education and educational-technology literature.

Method

This study employs a Design-Based Research (DBR) approach, a research methodology aimed at developing contextualised educational solutions through an iterative cycle of design, implementation, analysis, and refinement (Hoadley & Campos, 2022). The DBR method was chosen for its capacity to integrate theory and practice effectively, yielding practical solutions relevant to the needs of higher education—particularly in developing technology-based assessment tools. Specifically, the study follows the DBR stages as follows.

First, the stage of identifying and analysing the problem was carried out through direct classroom observation, in-depth interviews with lecturers and students, and an analysis of learning documents. This stage aimed to gain a thorough understanding of the challenges faced in assessing students' critical-thinking skills.

Second, the solution-development stage involved designing the Custom GPT model based on Bloom's Taxonomy, focusing on the abilities of synthesis (C4), evaluation (C5), and creation (C6). At this stage, GPT prompts were systematically composed and validated through expert judgement by specialists in educational technology and assessment.

Third, the iterative testing and refinement stage was conducted through several implementation cycles in real learning situations, involving students as direct users. Each cycle included the collection of qualitative and quantitative data to evaluate the effectiveness, practicality, and student engagement with the developed model. The test results were then used to continuously improve and refine the system.

Fourth, the reflection stage was undertaken to strengthen the theoretical and practical foundations of the developed solution. This process included an in-depth analysis of the collected data to evaluate the solution's impact on improving the quality of the learning process, particularly in supporting the development of critical-thinking skills.

Fifth, the documentation and dissemination stage comprised the preparation of a comprehensive research report and the publication of findings through scholarly journals, seminars, or other media to extend the benefits of this study to the academic community and educational practitioners.

This study employed purposive sampling with respondents selected from students in the Mathematics Education programme who were actively attending higher-education courses. Fourteen students were chosen as respondents, representing various levels of academic ability and experience with learning technology. This number of respondents was deemed sufficient to provide comprehensive and in-depth information on the effectiveness and practicality of the Custom GPT model in a real context. The respondents were directly involved in every testing stage and provided feedback through questionnaires and in-depth interviews to support the system's iterative refinement.

The evaluation instrument, consisting of GPT prompts, was validated using an expert-judgement method in which three experts in educational technology, educational assessment, and AI reviewed the instrument. The experts evaluated the clarity, relevance, and alignment of the prompts with the research objectives and the targeted thinking levels in Bloom's Taxonomy. The validation results were analysed using the Content Validity Index (CVI), with a minimum criterion of 0.80 to ensure high validity.

The evaluation procedure involved several stages. First, the respondents received clear and detailed instructions on using the Custom GPT system. Second, the respondents interacted with the system to complete critical-thinking tasks based on GPT prompts during a learning session. Third, the students' interactions with the system were evaluated using an assessment rubric based on indicators for synthesis, evaluation, and creation abilities. Fourth, the obtained data were analysed statistically and descriptively to assess system performance and provide qualitative and quantitative feedback for further refinement of the system.

By following this DBR cycle, the research is expected to produce a solution that is not only theoretical but also makes a tangible contribution to learning practice in higher education.

Result and Discussion

Identification and Analysis of the Problem

In the initial stage of this design-based study, a needs analysis was carried out to pinpoint specific issues in assessing critical-thinking skills among students of the Mathematics Education Study Program at Universitas Muhammadiyah Sorong. This problem-

analysis process comprised several steps namely, direct observation of learning activities, in-depth interviews with lecturers and students, and an analysis of learning documents.

Observations were conducted across six learning sessions in different classes, involving four lecturers whose courses emphasised the development of critical-thinking skills. The observations revealed that the critical-thinking assessment methods currently employed still rely on conventional instruments namely written questions and essay examinations, which are sub-optimal in providing prompt and detailed feedback to students. Moreover, the assessment process remains heavily dependent on lecturers' subjective interpretations, resulting in inconsistent outcomes across classes.

Subsequently, in-depth interviews were conducted with four lecturers and ten students from different semesters. These interviews revealed several challenges encountered in assessing critical-thinking skills, (1) Lecturers find it difficult to objectively assess students' critical-thinking processes in real time while instruction is underway. (2) Students are dissatisfied with the quality of feedback generated by current assessments, which dampens their motivation to improve their critical-thinking skills. (3) No existing assessment system offers in-depth analysis of specific dimensions synthesis (C4), evaluation (C5), and creation (C6), that constitute essential components of Bloom's Taxonomy.

An analysis of learning documents including Semester Learning Plans (RPS), assessment instruments, and learning-evaluation reports from the past two years, likewise revealed that most critical-thinking assessment instruments are still general in nature and fail to measure higher-order thinking aspects in depth. This has caused a gap between the intended learning objectives and students' actual attainment of critical-thinking skills.

Based on the results of the observations, interviews, and document analysis, it is clear that an urgent need exists to develop a new assessment instrument capable of effectively and objectively measuring students' critical-thinking abilities while also providing informative, in-depth, real-time feedback. Artificial-intelligence (AI) technology, particularly a customised GPT model, has been identified as a potential solution to address this issue. By adopting an AI-based approach, lecturers and students can directly monitor the development of students' critical-thinking skills and take timely corrective actions.

The findings from this needs-analysis stage form the basis for developing the Custom GPT model in the subsequent phase of this study, ensuring that the solution devised is genuinely relevant and effective in addressing the real problems encountered within the higher-education context.

Prompt Development and Testing

As part of developing the Custom GPT-based critical-thinking assessment system, six prompt iterations were designed, tested, and refined. Each iteration aimed to enhance the system's ability to assess higher-order thinking skills, particularly synthesis (C4), evaluation (C5), and creation (C6) as delineated in Bloom's Taxonomy. This iterative approach ensures that the system not only provides an interactive learning experience but also delivers accurate and relevant assessments.

Prompt 1 began with ChatGPT acting as a learning guide, facilitating critical questioning and in-depth discussion with students. In Prompt 2, this approach was enhanced by linking students' personal experiences to academic concepts, encouraging analysis and reflection. Prompts 3 through 5 introduced a more structured framework incorporating five stages—exploration, analysis, synthesis, evaluation, and creation, systematically designed to assess students' critical-thinking abilities.

Table 1. Prompt Development and Testing

Prompt Iteration	Description	Key Features
Prompt 1	ChatGPT acts as a learning guide, providing guidance through critical questions and in-depth discussions.	Initial guidance; critical questions and in-depth discussions.
Prompt 2	Connecting students' personal experiences to academic concepts to facilitate analysis and reflection.	Linking personal experiences to academic concepts; encouraging reflection.
Prompt 3	A more structured framework with five stages: exploration, analysis, synthesis, evaluation, and creation.	Organized five-stage structure; delving into critical thinking skills.
Prompt 4	Expanding the stages of analysis, evaluation, and creativity by integrating students' real-world experiences.	Integration of real-world experiences for deeper analysis and evaluation.
Prompt 5	Strengthening the previous five stages with a focus on systematically testing higher-order critical thinking skills.	Focus on systematic testing of higher-order thinking skills.

Prompt Iteration	Description	Key Features
Prompt 6	Refining the five stages: starting with clarification, ending with assessing the quality of responses in C4, C5, and C6 domains. Emphasis on synthesis, evaluation, and creation.	Assessment of response quality in C4 (Synthesis), C5 (Evaluation), C6 (Creation); comprehensive learning process.

The final iteration, Prompt 6, represents the culmination of improvements derived from the earlier tests. This prompt comprises five stages that start with clarification through students' personal experiences, followed by encouraging analysis, synthesis new ideas, evaluating arguments, and culminating in the creation of new solutions or projects based on their understanding. At the closing stage, students are asked to summarise their learning, and the system provides an assessment based on the quality of responses in each domain (C4, C5, C6). Compared with the previous iterations, this prompt demonstrates superiority in comprehensively facilitating higher-order thinking processes, as detailed in Table 1.

The results of this iterative development highlight the benefits of identifying and strengthening weaknesses in the initial design. With the final structured prompt, the Custom GPT system not only effectively evaluates students' critical-thinking skills but also provides detailed feedback to support the development of their analytical and creative abilities. These findings reinforce the potential for wider implementation of the system in higher-education contexts.

Testing and Refinement

The Custom GPT system was tested with fourteen respondents and evaluated across three domains of higher-order thinking skills: Synthesis (C4), Evaluation (C5), and Creation (C6). The results, summarised in Figure 3, reveal variations in individual performance, with each domain's scores as well as the overall score depicted in the graph.

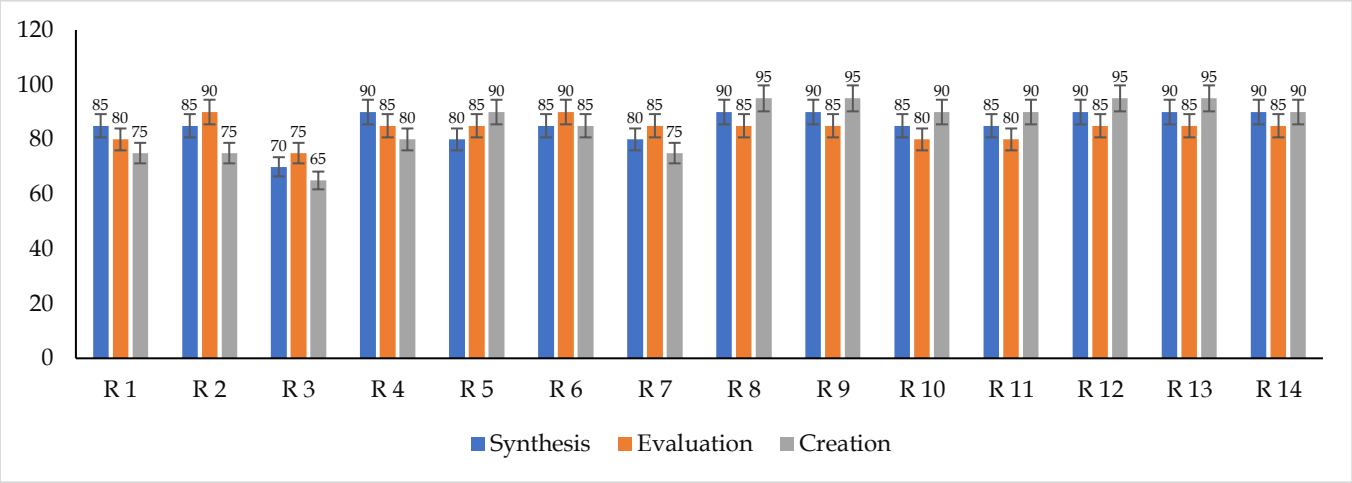

Figure 1. Distribution of Synthesis, Evaluation, and Creation Scores

The graph indicates that most respondents performed well in the Synthesis domain, with the highest score of 90 achieved by several respondents, including Respondent 8 and Respondent 9. Evaluation scores exhibited a similar trend, with a maximum score of 90 also attained by several participants. However, in the Creation domain, there was greater variation. While Respondent 8 and Respondent 9 also achieved the highest score (95) in this domain, others, such as Respondent 3, recorded lower scores, with 65 being the lowest.

The final scores show that the majority of respondents fell within the Good to Very Good categories, with the highest overall score of 90 achieved by several respondents, while the lowest score of 70 was recorded by Respondent 3. The graph also highlights a relative balance in the Synthesis and Evaluation domains, whereas the Creation domain presents opportunities for further improvement for most respondents.

The analysis of these results demonstrates that the Custom GPT system successfully measured higher-order thinking skills, particularly in the Synthesis and Evaluation domains (Wang et al., 2025). However, the variation in scores for the Creation domain suggests the need to strengthen strategies for fostering students' innovative and creative abilities. Based on observations, the system could be enhanced by incorporating more

real-world, application-based problem-solving scenarios to improve Creation skills.

The graph provides a clear visual representation of respondents' strengths and weaknesses in each domain, supporting detailed analysis and recommendations for system refinement. This finding underscores the potential of the Custom GPT-based approach as an effective tool for evaluating higher-order thinking skills in higher education (Jelica et al., 2023; Naznin et al., 2025; Wu et al., 2025).

Practicality Analysis

As part of the evaluation, a practicality analysis of the Custom GPT system was conducted based on the criteria of Ease of Use, Engagement, and Efficiency. This process utilized a formula adapted from (Akbar, 2013).

Table 2. Summary of the categories and evaluation

Category	Percentage (Vp) %	Criteria
Ease of Use	86.70	Very Practical
Engagement	81.80	Practical
Efficiency	85.10	Very Practical

The analysis results indicate that Ease of Use scored 86.7%, falling into the Very Practical category. This demonstrates that the system is easy to operate for users, including both students and lecturers, without significant difficulties. Meanwhile, Engagement scored 81.8%, categorized as Practical. This result reflects that the system sufficiently captures user attention, though there are opportunities to improve this aspect for better optimization. Additionally, Efficiency scored 85.1%, also classified as Very Practical, indicating that the system can be effectively used to complete tasks quickly and efficiently.

Overall, the practicality analysis results show that the Custom GPT system has a high level of practicality. Two out of three categories fall into the Very Practical category, reinforcing the feasibility of implementing this system in higher education. However, improvements in the engagement aspect could enhance the overall user experience and support better involvement. With these results, the Custom GPT system holds significant potential for broader implementation.

Reflection & Theory Strengthening

The development and testing process of the Custom GPT system provided important insights that reinforce Bloom's Taxonomy theory in the assessment of critical thinking skills. Bloom's Taxonomy, known as a systematic framework for measuring higher-order thinking skills (C4-Synthesis, C5-Evaluation, C6-Creation), is highly relevant in supporting critical skills-based learning in the digital era (Essien et al., 2024). The iterative approach used in this study, through six prompt iterations, demonstrated how artificial intelligence can be effectively integrated to produce accurate and measurable evaluations. For instance, testing on respondents showed that the system could identify strengths in Synthesis and Evaluation, while providing critical insights into challenges within the Creation domain, aligning with the literature that creation requires more contextual and complex stimulation (Surin & Damrongpanit, 2024).

Additionally, the system's practicality analysis reflects the relevance of AI-based technology in enhancing the efficiency and usability of learning tools. Bitzenbauer (2023) noted that AI technology, including generative models like ChatGPT, holds great potential for delivering personalized feedback, supporting students' self-reflection and improvement. The findings of this study align with those results, as respondents rated the system as "Very Practical" in terms of ease of use (86.7%) and efficiency (85.1%). However, the engagement score (81.8%), categorized as "Practical," highlights the need to improve the system's interface design and interactivity, supporting (Elen & Verburgh, 2023) recommendation on the importance of designs that encourage student engagement.

This reflection indicates that the Custom GPT system functions not only as an evaluation tool but also as a learning medium that supports the development of critical thinking skills, in line with constructivist theory that learning must involve real-world experiences (Vygotsky & Cole, 1978). Further adjustments to the engagement aspect, such as integrating real-world application-based problem-solving scenarios, could enhance the system's effectiveness. Thus, this study strengthens the theoretical framework of education based on Bloom's Taxonomy while proving that artificial intelligence can be innovatively used to support data-driven assessment and learning in higher education.

Conclusion

This study has successfully developed and tested a Custom GPT model designed to assess critical thinking skills in higher education students. The system comprehensively evaluated the higher-order thinking skills of synthesis (C4), evaluation (C5), and creation (C6) based on Bloom's Taxonomy. Using an iterative design-based research (DBR) methodology, the study demonstrated that the Custom GPT system is capable of providing accurate and objective assessments while delivering real-time feedback, which is critical for fostering students' analytical and creative abilities. The findings address the primary research question by proving that the integration of artificial intelligence (AI)

with educational assessment can effectively measure critical thinking skills. The results revealed high practicality scores, with Ease of Use and Efficiency categorized as "Very Practical," and Engagement as "Practical." These outcomes indicate that the system is not only effective but also user-friendly and adaptable to educational contexts. In addition to its practical contributions, the study strengthens theoretical frameworks by integrating AI-driven evaluation with Bloom's Taxonomy. The system bridges existing gaps in traditional assessment methods by offering a data-driven, adaptive, and scalable solution for higher education institutions. This research paves the way for further exploration into the role of AI in enhancing both learning and assessment, with potential applications in diverse educational and professional settings.

Acknowledgments

The authors extend their heartfelt appreciation to the Council for Higher Education, Research, and Development (Diktilitbang) of the Muhammadiyah Central Board for administering the RisetMu National Research Grant (Batch VIII) that underpinned this study. We are equally grateful to the Institute for Research and Community Service (LP3M) of Universitas Muhammadiyah Sorong for administrative and logistical support throughout the project.

Author Contributions

Conceptualization, M.F. and F.E.M.; methodology, M.F.; software, T.H.I.A.; validation, M.F., F.E.M. and M.S.K.; formal analysis, M.S.K.; investigation, K.P.A.; resources, T.H.I.A.; data curation, M.F.; writing—original draft preparation, M.F.; writing—review and editing, F.E.M.; visualization, K.P.A.; supervision, T.H.I.A.; project administration, M.F. All authors have read and agreed to the published version of the manuscript.

Funding

This work was funded by the Council for Higher Education, Research, and Development (Diktilitbang) of the Muhammadiyah Central Board through the RisetMu National Research Grant, Batch VIII – Fundamental Research Scheme II (Contract No. 0258.101/I.3/D/2025). The Council also covered the article-processing charge (APC). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

Akbar, S. (2013). *Instrumen Perangkat Pembelajaran*. Bandung: PT Remaja Rosdakarya.

Andiopenta. (2024). Use of Artificial Intelligence (AI) in Writing Scientific Works. *Jurnal Penelitian Pendidikan IPA*, *10*(12), 10196–203. https://doi.org/10.29303/jppipa.v10i12.9414

Anisah, D. C., Nasrudin, H., & Sukarmin. (2025). Profile of Critical Thinking Skills of Phase F Learning in Chemistry Subjects Acid-Base Material. *Jurnal Penelitian Pendidikan IPA*, *11*(1), 17–27. https://doi.org/10.29303/jppipa.v11i1.9842

Bitzenbauer, P. (2023). ChatGPT in Physics Education: A Pilot Study on Easy-to-Implement Activities. *Contemporary Educational Technology*, *15*(3). https://doi.org/10.30935/cedtech/13176

Cahyani, C. P. N., & Setiawan, E. P. (2024). How Motivation Mediated Science Achievement of Indonesian Students: A Path Analysis on PISA 2018 Data. *Jurnal Penelitian Pendidikan IPA*, *10*(12), 10493–501. https://doi.org/10.29303/jppipa.v10i12.10012

Elen, J., & Verburgh, A. (2023). Fostering Critical Thinking: Features of Powerful Learning Environments. *European Journal of Education*, *58*(3), 434–46. https://doi.org/10.1111/ejed.12568

Essien, A., Bukoye, O. T., O'Dea, X., & Kremantzis, M. (2024). The Influence of AI Text Generators on Critical Thinking Skills in UK Business Schools. *Studies in Higher Education*, *49*(5), 865–82. https://doi.org/10.1080/03075079.2024.2316881

Hoadley, C., & Campos, F. C. (2022). Design-Based Research: What It Is and Why It Matters to Studying Online Learning. *Educational Psychologist*, *57*(3), 207–20. https://doi.org/10.1080/00461520.2022.2079128

Hopfenbeck, T. N., Zhang, Z., Sun, S. Z., Robertson, P., & McGrane, J. A. (2023). Challenges and Opportunities for Classroom-Based Formative Assessment and AI: A Perspective Article. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1270700

Indrašienė, V., Jegelevičienė, V., Merfeldaitė, O., Penkauskienė, D., Pivorienė, J., Railienė, A., & Sadauskas, J. (2023). Value of Critical Thinking in the Labour Market: Variations in Employers' and Employees' Views. *Social Sciences*, *12*(4). https://doi.org/10.3390/socsci12040221

Jelica, S., Miroslav, M., & Uroš, Š. (2023). Possible applications of ChatGPT in higher education. *International Conference on Technological Ecosystems for Enhancing Multiculturality*, 135–144. https://doi.org/10.1007/978-981-97-1814-6_13

Kadrija, R., Shatri, Z. G., & Këndusi, V. V. (2022). Effects of Critical Thinking Implementation on Enhancing of Teaching Quality. *Journal of Educational and Social Research*, *12*(6), 236–45. https://doi.org/10.36941/jesr-2022-0159

Kofahi, A., Mohammed, & Husain, A. (2025). ChatGPT for Operating Systems: Higher-Order Thinking in

Focus. *Journal of Information Technology Education: Research*, *24*(001). https://doi.org/10.28945/5382

Lee, H.-Y., Chen, P.-H., Wang, W.-S., Huang, Y.-M., & Wu, T.-T. (2024). Empowering ChatGPT with Guidance Mechanism in Blended Learning: Effect of Self-Regulated Learning, Higher-Order Thinking Skills, and Knowledge Construction. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00447-4

López, F., Contreras, M., Nussbaum, M., Paredes, R., Gelerstein, D., Alvares, D., & Chiuminatto, P. (2023). Developing Critical Thinking in Technical and Vocational Education and Training. *Education Sciences*, *13*(6). https://doi.org/10.3390/educsci13060590

Naznin, K., Al Mahmud, A., Nguyen, M. T., & Chua, C. (2025). ChatGPT integration in higher education for personalized learning, academic writing, and coding tasks: a systematic review. *Computers*, *14*(2), 53. https://doi.org/10.3390/computers14020053

OECD. (2024). PISA 2022 Results (Volume I): The State of Learning and Equity in Education. In *Perfiles Educativos* (Vol. 46, Issue 183). OECD Publishing. https://doi.org/10.22201/iisue.24486167e.2024.183.61714

Surin, S., & Damrongpanit, S. (2024). Developing Students' Critical Thinking: Examining the Influence of Learning Management Approaches Through Meta-Analysis and Propensity Score Matching. *European Journal of Educational Research*, *13*(3), 1391–1409. https://doi.org/10.12973/eu-jer.13.3.1391

Swindell, A., Greeley, L., Farag, A., & Verdone, B. (2024). Against Artificial Education: Towards an Ethical Framework for Generative Artificial Intelligence (AI) Use in Education. *Online Learning*, *28*(2). https://doi.org/10.24059/olj.v28i2.4438

van den Berg, G., & du Plessis, E. (2023). ChatGPT and Generative AI: Possibilities for Its Contribution to Lesson Planning, Critical Thinking and Openness in Teacher Education. *Education Sciences*, *13*(10). https://doi.org/10.3390/educsci13100998

Vygotsky, L. S., & Cole, M. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.

Wang, W.-S., Lin, C.-J., Lee, H.-Y., Huang, Y.-M., & Wu, T.-T. (2025). Enhancing self-regulated learning and higher-order thinking skills in virtual reality: the impact of ChatGPT-integrated feedback aids. *Education and Information Technologies*, 1–27. https://doi.org/10.1007/s10639-025-13557-x

Wu, T.-T., Lee, H.-Y., Chen, P.-H., Lin, C.-J., & Huang, Y.-M. (2025). Integrating peer assessment cycle into ChatGPT for STEM education: A randomised controlled trial on knowledge, skills, and attitudes enhancement. *Journal of Computer Assisted Learning*, *41*(1), 13085. https://doi.org/10.1111/jcal.13085

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A Scoping Review on How Generative Artificial Intelligence Transforms Assessment in Higher Education. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00468-z.

Xiaolei, S., & Teng, M. F. (2024). Three-Wave Cross-Lagged Model on the Correlations between Critical Thinking Skills, Self-Directed Learning Competency and AI-Assisted Writing. *Thinking Skills and Creativity*, *52*. https://doi.org/10.1016/j.tsc.2024.101524