Jurnal Aksara
https://aksara.kemdikbud.go.id/index.php/aksara
ISSN: 0854-3283 (print)
ISSN: 2580-0353 (online)

THE ROLE OF CRITICAL THINKING IN THE PREPARATION OF HIGH-LEVEL INDONESIAN QUESTIONS BY PRE-SERVICE TEACHERS

Peran Berpikir Kritis dalam Penyusunan Soal Bahasa Indonesia Tingkat Tinggi oleh Guru Prajabatan

Sri Wahyuni[a]*, Frida Siswiyanti[b], Itznaniyah Umie Murniatie[c], Prayitno Tri Laksano[d], Lilik Wahyuni[e], Ifit Novita Sari[f]
[abcdf]Universitas Islam Malang
Jalan Mayjen Haryono No. 193, Dinoyo, Lowokwaru, Malang 65144, Jawa Timur, Indonesia.
[e]Universitas Brawijaya
Jalan Veteran, Malang 65145, Jawa Timur, Indonesia.
Pos-el: sriwy@unisma.ac.id, fridasiswiyanti@unisma.ac.id, itznaniyahmurniatie@unisma.ac.id, prayitno27@unisma.ac.id, lilikwahyuni@ub.ac.id, inovsari@unisma.ac.id

Abstract
This study aims to (1) describe the role of pre-service teachers' critical thinking skills in their ability to formulate higher-order Indonesian language questions, and (2) identify which aspects of critical thinking have the most significant influence on this ability. A quantitative correlational approach was employed, with data collected from 71 pre-service teachers through a critical thinking skills test and an analysis of the higher-order questions they created. The data were analyzed using correlation tests, one-way ANOVA, and Post Hoc LSD analysis to examine relationships and differences in performance based on levels of critical thinking. The findings show a significant positive relationship between critical thinking and the ability to design analytical, evaluative, and creative questions. Pre-service teachers with higher levels of critical thinking demonstrated better performance in constructing cognitively complex questions. The ANOVA results ($F = 55.632$; $p < 0.05$) confirmed significant differences among the low-, medium-, and high-critical-thinking groups. These findings affirm that critical thinking is a key foundation for developing higher-order questioning skills. Therefore, teacher education programs should integrate inquiry-based learning, metacognitive reflection, and technology-based supports such as AI-assisted scaffolding to strengthen pre-service teachers' critical and reflective thinking abilities.
Keywords: *critical thinking, higher-order questions, Indonesian Language, pre-service teachers', teacher professional education*

Abstrak
Penelitian ini bertujuan untuk (1) mendeskripsikan peran keterampilan berpikir kritis guru prajabatan dalam kemampuan mereka menyusun pertanyaan Bahasa Indonesia tingkat tinggi, serta (2) mengidentifikasi aspek berpikir kritis yang paling berpengaruh terhadap kemampuan tersebut. Penelitian ini menggunakan pendekatan kuantitatif dengan desain korelasional. Data dikumpulkan dari 71 guru prajabatan melalui tes berpikir kritis dan analisis terhadap pertanyaan tingkat tinggi yang mereka buat. Data dianalisis menggunakan uji korelasi, ANOVA satu arah, dan uji Post Hoc LSD untuk melihat hubungan dan perbedaan kemampuan berdasarkan tingkat berpikir kritis. Hasil penelitian menunjukkan adanya hubungan positif yang signifikan antara keterampilan berpikir kritis dan kemampuan menyusun pertanyaan analitis, evaluatif, serta kreatif. Guru prajabatan dengan kemampuan berpikir kritis tinggi menunjukkan hasil yang lebih baik dalam menyusun pertanyaan kompleks secara kognitif. Hasil ANOVA ($F = 55{,}632$; $p < 0{,}05$) juga menegaskan adanya perbedaan yang bermakna antara kelompok berpikir kritis rendah, sedang, dan tinggi. Temuan ini menegaskan bahwa berpikir kritis merupakan dasar penting dalam mengembangkan kemampuan bertanya tingkat tinggi. Oleh karena itu, program pendidikan guru perlu mengintegrasikan pembelajaran berbasis inkuiri, refleksi metakognitif, serta dukungan teknologi seperti AI-assisted scaffolding untuk memperkuat kemampuan berpikir kritis dan reflektif calon guru.

Kata-kata kunci: berpikir kritis, pertanyaan tingkat tinggi, Bahasa Indonesia, guru prajabatan, pendidikan profesi guru

Informasi Artikel

Naskah Diterima	Naskah Direvisi akhir	Naskah Diterbitkan
24 Juni 2025	10 November 2025	12 Desember 2025

Cara Mengutip

Wahyuni, Sri., dkk. (2025). The Role of Critical Thinking in The Preparation of High-Level Indonesian Questions by Pre-Service Teachers. *Aksara. 37*(2). 454-463. http://dx.doi.org/10.29255/aksara.v37i2.4888.454-463

INTRODUCTION

In the rapidly changing educational landscape of the 21st century, critical thinking (CT) is widely recognized as a cornerstone of effective teacher education. Teachers are expected not only to master content knowledge but also to cultivate students' higher-order thinking through meaningful and cognitively demanding learning activities. Within language education, the ability to design high-level questions—those that require analysis, synthesis, and evaluation—plays a strategic role in promoting deep learning and critical engagement (Chan et al., 2002; Ilhan & Gezer, 2017; Vuleta & Martinović, 2019). However, empirical evidence indicates that pre-service teachers frequently struggle to formulate such questions, often relying on lower-order, factual prompts (Lu & Zheng, 2024).

The researchers' interest in this study emerged from observations that, despite the Indonesian Merdeka Curriculum emphasizing higher-order thinking skills (HOTS), many pre-service teachers still exhibit difficulties in designing assessment items that foster critical reasoning. This phenomenon reflects a broader issue in teacher preparation programs: the insufficient integration between CT training and practical question-design skills. As Orakcı & Khalili (2025) emphasize, CT dispositions such as open-mindedness, systematicity, and truth-seeking are foundational for constructing cognitively complex questions, yet these dispositions remain unevenly developed among teacher candidates.

Recent studies have reaffirmed the critical role of metacognitive awareness and self-efficacy in mediating the relationship between CT and question formulation (Erdoğan & Kalkan, 2024; Küçükaydın, 2024). Pre-service teachers with higher metacognitive control demonstrate greater ability to plan, monitor, and evaluate their own thinking processes when creating high-level questions. Nevertheless, research in Indonesia shows that most teacher candidates exhibit only moderate CT dispositions, with persistent challenges in reflective and analytical questioning (Fitriani et al., 2018).

Empirical investigations further confirm that strengthening CT through inquiry-based and argument-driven instructional models (Metin Peten, 2022; Saputri et al., 2020) significantly enhances teachers' questioning competence. Similarly, the integration of AI-supported learning environments has emerged as a promising innovation for developing both CT and questioning skills. Large language models (LLMs), when ethically and contextually applied, can assist pre-service teachers in generating and evaluating high-level language questions (Lai et al., 2025; Lodovico Molina et al., 2024; Scaria et al., 2024). However, overreliance on automated tools without sufficient pedagogical grounding may reduce the reflective depth of CT processes, necessitating careful integration and teacher agency (Mondal, 2025; Moorhouse et al., 2025).

Despite these advances, there remains a paucity of empirical research explicitly examining how CT skills influence pre-service teachers' ability to formulate higher-order Indonesian language questions. While prior studies have explored this relationship in EFL or

STEM education contexts (Wulandari et al., 2021; Yıldız-Feyzioğlu & Kıran, 2022), few have focused on Indonesian language teacher education-a gap that this research seeks to address. Accordingly, this study aims to:

1. describe the role of pre-service teachers' critical thinking skills in their ability to formulate high-level Indonesian language questions; and
2. identify which dimensions of CT most strongly predict the ability to develop higher-order question types.

Understanding this relationship is crucial for designing targeted, evidence-based teacher education programs that strengthen both critical reasoning and question-design competence. Furthermore, this study contributes conceptually and empirically by bridging CT theory, higher-order questioning practices, and the contextual realities of Indonesian language instruction within the national teacher education framework.

METHOD

This study employed a quantitative method using a correlational research design. This design was selected to obtain a comprehensive understanding of the relationship between pre-service teachers' critical thinking skills and their ability to formulate higher-order questions in Bahasa Indonesia instruction. Data were collected through a critical thinking skills test and an assessment of the ability to construct higher-order questions.

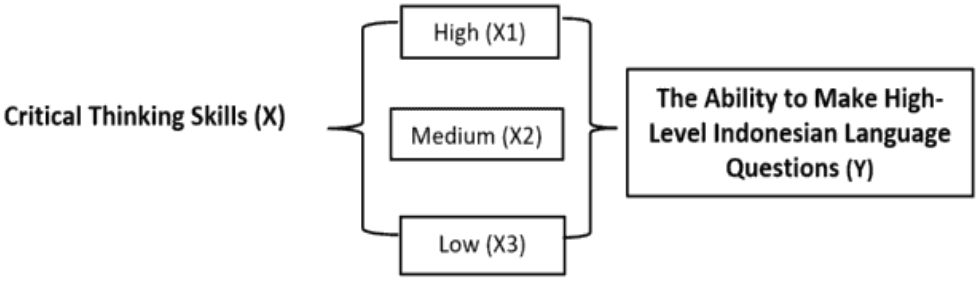

Figure 1. Research Design

The population of this study comprised pre-service teachers enrolled in the Teacher Professional Education (Pendidikan Profesi Guru/PPG) program at Universitas Islam Malang in 2024. The sample consisted of 71 pre-service teachers from the Indonesian Language Education Study Program, selected through purposive sampling. This sampling technique was based on participants' engagement in courses related to question construction and critical thinking skills.

The instruments used in this study included a critical thinking skills test adapted from (Ennis, 2018) critical thinking framework and an assessment rubric for higher-order question construction based on the revised Bloom's Taxonomy. Data were collected through the administration of the critical thinking test and a task requiring participants to generate higher-order questions. Quantitative data were analyzed using a one-way ANOVA to examine differences in the ability to construct higher-order questions across levels of critical thinking skills (high, moderate, and low).

The validity of the quantitative instruments was established through content validation by experts in language education. Reliability was measured using Cronbach's alpha coefficient. For the qualitative data, the credibility of interview findings was ensured through source triangulation and member checking. This study adhered to ethical research standards by obtaining written informed consent from participants prior to data collection, maintaining the confidentiality of participants' identities, and ensuring voluntary participation without coercion. A limitation of this study is the restricted sample size, which was confined to one university.

RESULT AND DISCUSSION

Results

This section presents an investigation into whether critical thinking skills contribute to the development of pre-service teachers' proficiency in constructing higher-order thinking (HOT) questions. The data analysis procedures and findings are described as follows. The results of the descriptive analysis are presented below.

Table 1. Statistics Descriptive Critical Thinking Skills

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Critical Thinking Skills	71	27.00	73.00	100.00	87.0563	8.43443
Valid N (listwise)	71					

The number of participants (N) in this study was 71. The range of Critical Thinking Skills scores varied from 73.00 to 100.00, yielding a total range of 27.00. The mean score for Critical Thinking Skills was 87.0563, indicating that, in general, participants demonstrated a relatively high level of critical thinking ability. The standard deviation of 8.43443 reflects a moderate degree of variability among participants' scores, suggesting that although the average score was high, some participants scored notably above or below the mean.

Table 2. Classification of Critical Thinking Skills

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low	6	8.5	8.5	8.5
	Medium	54	76.1	76.1	84.5
	High	11	15.5	15.5	100.0
	Total	71	100.0	100.0	

Of the 71 participants, 6 individuals (8.5%) were categorized as having low critical thinking skills. The majority of participants, 54 individuals (76.1%), fell into the medium category, while 11 participants (15.5%) demonstrated high critical thinking skills.

The data in the two tables above indicate that the majority of pre-service teachers possess critical thinking skills within the medium category (76.1%). A small proportion (8.5%) displayed low critical thinking ability, highlighting a need to enhance such skills within this group. The relatively high average score for critical thinking (M = 87.06) suggests that, overall, the pre-service teachers in this study possess a solid cognitive foundation for developing higher-order thinking questions in Bahasa Indonesia instruction. The moderate score variability (SD = 8.43) indicates that, while differences in critical thinking ability exist among participants, these differences are not extreme.

Table 3. Descriptive Statistics of Higher-Order Questioning Ability

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Ability to Create HOTS Questions	71	36.33	56.67	93.00	84.4834	4.65940
Valid N (listwise)	71					

Number of Participants (N): A total of 71 individuals participated in this study. Score Range: The scores for higher-order question development ranged from 56.67 to 93.00, with a range of 36.33. Minimum Score: The lowest score obtained by a participant was 56.67. Maximum Score: The highest score obtained was 93.00. Mean Score: The average score for higher-order question development was 84.48, indicating that, overall, participants

demonstrated a relatively high level of ability in constructing higher-order thinking questions. Standard Deviation: The standard deviation was 4.66, indicating relatively low variability among participants' scores. This suggests that most scores clustered closely around the mean, reflecting consistent performance in higher-order question development.

The findings presented in the table above reveal that participants exhibited a relatively high level of ability in developing higher-order thinking questions, with a mean score of 84.48. This suggests that the pre-service teachers involved in this study generally possess strong skills in this area. The low variability in scores, as indicated by a standard deviation of 4.66, shows that differences in ability among participants were not substantial, implying a relatively uniform level of competence. Despite the generally high performance, the wide score range (36.33) indicates the presence of a few participants who may require additional support to improve their skills in developing higher-order questions-particularly those whose scores are closer to the minimum value (56.67). While the overall outcomes are promising, the existence of lower-performing participants highlights the need for continued reinforcement in this area. Targeted training or mentoring programs could be beneficial, particularly for those scoring below the average, to further enhance their ability to construct effective higher-order thinking questions.

Table 4. ANOVA Test on Critical Thinking Skills

Critical Thinking Skills

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3090.793	2	1545.397	55.632	.000
Within Groups	1888.981	68	27.779		
Total	4979.775	70			

The table above presents the results of the ANOVA (Analysis of Variance) test for critical thinking skills. The interpretation of each component is as follows:

Sum of Squares: Between Groups: The value of 3090.793 represents the amount of variability between the groups being compared, reflecting the differences in critical thinking scores across groups. Within Groups: The value of 1888.981 captures the variability within each group, representing differences in critical thinking scores among participants within the same group. Total: The total sum of squares is 4979.775, which is the sum of both between-group and within-group variances.

Degrees of Freedom (df): Between Groups: The degrees of freedom between groups is 2, indicating the number of groups compared minus one. Within Groups: The within-group degrees of freedom is 68, calculated by subtracting the number of groups from the total number of participants. Total: The total degrees of freedom is 70.

Mean Square: Between Groups: The mean square value of 1545.397 is obtained by dividing the between-group sum of squares by its degrees of freedom (3090.793 ÷ 2). Within Groups: The within-group mean square value of 27.779 is calculated by dividing the within-group sum of squares by its degrees of freedom (1888.981 ÷ 68).

F-Value: The F-statistic of 55.632 represents the ratio of between-group variance to within-group variance. A higher F-value indicates a greater difference between groups relative to the variation within groups.

Significance (Sig.): The significance value of 0.000 ($p < 0.05$) indicates that the differences between groups are statistically significant. In other words, there are meaningful differences in critical thinking skills among the groups analyzed.

Interpretation: The ANOVA results indicate a statistically significant difference in critical thinking skills across the groups ($p = 0.000$). This finding confirms that the groups under comparison exhibit significantly different levels of critical thinking skills. A substantial portion of the variance in critical thinking skills stems from differences between groups (3090.793),

which is much greater than the within-group variance (1888.981). The high F-value (55.632) suggests that critical thinking skills have a significant influence on the ability to construct higher-order thinking questions, supporting the hypothesis that this skill plays an important role in the process.

Table 5. Post Hoc Tests

Dependent Variable: Critical Thinking Skills LSD

(I) Category	(J) Category	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Low	Medium	-12.98148*	2.26810	.000	-17.5074	-8.4556
	High	-27.00000*	2.67493	.000	-32.3377	-21.6623
Medium	Low	12.98148*	2.26810	.000	8.4556	17.5074
	High	-14.01852*	1.74350	.000	-17.4976	-10.5394
High	Low	27.00000*	2.67493	.000	21.6623	32.3377
	Medium	14.01852*	1.74350	.000	10.5394	17.4976

*. The mean difference is significant at the 0.05 level.

This table presents the results of a Post Hoc test using the Least Significant Difference (LSD) method to compare pairs of critical thinking skill groups. The detailed explanation is as follows: Mean Difference Between Groups: The mean difference in critical thinking skills between the Low and Medium groups is -12.98148, indicating that the Low group scored lower than the Medium group. The mean difference between the Low and High groups is -27.00000, revealing that the Low group scored significantly lower than the High group. The Medium group scored 12.98148 points higher than the Low group and 14.01852 points lower than the High group. The High group demonstrated significantly higher scores than both the Low group (27.00000) and the Medium group (14.01852).

All pairwise group comparisons yielded p-values of 0.000, indicating that the differences between all group pairs are statistically significant at the 0.05 significance level. 95% Confidence Intervals: The confidence intervals do not include zero, confirming the significance of the mean differences between groups. For instance, the difference between the Low and Medium groups ranges from -17.5074 to -8.4556, reflecting a clear distinction between the two groups.

Interpretation: There are statistically significant differences in critical thinking skills among the Low, Medium, and High groups. The results demonstrate a tiered scoring pattern, where participants in the Low critical thinking group consistently performed worse than those in the Medium and High groups. These findings suggest a strong relationship between higher levels of critical thinking skills and greater ability in developing higher-order thinking questions, thereby supporting the hypothesis that critical thinking skills significantly influence performance in constructing such questions.

Discussion

The Significance of Critical Thinking in Teacher Education

The results of this study confirm that critical thinking (CT) skills have a significant positive correlation with pre-service teachers' ability to construct higher-order thinking (HOT) questions in Indonesian language learning. This finding aligns with the theoretical premise that CT underpins the formulation of analytical and evaluative questions that stimulate students' cognitive engagement (Facione, 2011; Halpern & Dunn, 2023).

From a teacher-education perspective, CT serves as a metacognitive engine that enables teachers to deconstruct problems, interpret learning contexts, and formulate questions that promote inquiry (Chan et al., 2002; Vuleta & Martinović, 2019). Consistent with Huang & Sang (2023), CT should be viewed not only as a cognitive disposition but also as a pedagogical competence essential for reflective decision-making in instruction.

In this study, pre-service teachers who demonstrated higher CT scores also produced more conceptually complex questions. This finding resonates with Yıldız-Feyzioğlu & Kıran (2022), who found that CT mediates the relationship between self-efficacy and the ability to reason argumentatively. Hence, CT acts as both a predictor and facilitator of advanced questioning ability.

Relationship Between Critical Thinking Levels and Question Formulation

Quantitative analysis revealed a significant difference across groups of critical-thinking proficiency ($F = 55.632$; $p < 0.05$). The Post Hoc results indicate that the high-CT group significantly outperformed both medium-and low-CT groups in developing HOT questions. This suggests that CT exerts a direct influence on question-design capability-a relationship also emphasized in Ilhan & Gezer (2017), who reported that mastery of higher cognitive taxonomies correlates with the complexity of assessment items.

The finding strengthens empirical patterns identified in Metin Peten (2022) and Saputri et al. (2020), where argument-driven inquiry and QASEE learning models substantially enhanced CT and question-formulation skills. These interventions explicitly trained participants to connect reasoning steps with Bloom's taxonomy-similar to the way higher-CT participants in the present study structured questions that demand analysis and synthesis.

Thus, the data affirm that CT competence predicts both the quality and depth of pedagogical questioning, reinforcing earlier research linking reasoning dispositions to cognitive challenge (Wulandari et al., 2021; Zhang et al., 2024).

Pedagogical Implications: Developing Reflective and Analytical Questioning Skills

The finding that most participants fell within the medium CT category (76.1 %) underscores an urgent need for systematic CT scaffolding in teacher-education curricula. According to Orakcı & Khalili (2025), integrating reflective dialogue and self-efficacy training significantly enhances teachers' cognitive flexibility and critical questioning ability.
Pedagogical programs should therefore embed explicit CT components-argumentation, metacognitive monitoring, and inquiry tasks-into language-teaching methodology courses. Erdoğan & Kalkan (2024) demonstrate that metacognitive awareness predicts improvement in CT, suggesting that reflective self-assessment after question construction could strengthen both awareness and skill transfer.

Furthermore, the QASEE model (Saputri et al., 2020) and AI-supported scaffolding environments (Lai et al., 2025; Luo et al., 2025) illustrate how digital platforms can facilitate iterative feedback and guided question generation, leading to enhanced HOT question quality. Such integration is particularly relevant for Indonesian pre-service teachers adapting to blended-learning contexts.

Integrating AI-Supported Tools and Inquiry-Based Pedagogies

Recent Scopus-indexed research indicates that AI-enhanced environments provide new opportunities for fostering CT and questioning competence. For example, Luo et al. (2025) found that procedural scaffolding within human-AI collaborative settings improved teachers' questioning accuracy and diversity. Xiao et al. (2025) confirmed that fine-tuned large language models (LLMs) can assess and generate higher-order questions across Bloom's taxonomy domains.

While the present study did not employ AI tools directly, its findings align with these insights by emphasizing cognitive depth and structured reasoning in question formulation. The implication is that AI-supported training could complement conventional CT development by

providing real-time analytical feedback while maintaining human reflection as the core of teacher learning (Mondal, 2025; Moorhouse et al., 2025).

Contextual and Cultural Considerations in Critical Thinking Training

It is important to note that CT and questioning behaviors are context-dependent and influenced by sociocultural values. The relatively homogeneous results among Indonesian pre-service teachers may stem from shared educational experiences emphasizing convergent, exam-oriented learning. As highlighted by Santos Meneses (2020), CT development requires pedagogical approaches sensitive to cultural norms and classroom discourse styles.

Therefore, teacher-education programs in Indonesia should adopt contextually responsive CT instruction-linking reflective inquiry with local linguistic and cultural frameworks. Incorporating local discourse practices, such as narrative reasoning and communal dialogue, may help bridge global CT paradigms with indigenous educational values.

Limitations and Directions for Future Research

This study was limited by its sample size (n = 71) and confinement to a single institution, restricting the generalizability of results. Future research could employ multi-institutional or longitudinal designs to examine how CT training influences questioning ability over time. Moreover, as suggested by Huang & Sang (2023) and Scaria et al. (2024), mixed-method approaches combining automated question-analysis tools with human expert judgment would yield more comprehensive evaluations of question quality.

In summary, the findings corroborate a significant and positive linkage between critical thinking and higher-order question formulation among pre-service Indonesian language teachers. Strengthening CT through reflective, inquiry-based, and AI-assisted pedagogies is pivotal for equipping future teachers with the analytical tools necessary to design cognitively engaging language learning experiences.

By situating CT within both cognitive theory and cultural practice, this research contributes to the broader goal of developing teacher education that is evidence-based, technologically adaptive, and contextually grounded.

CONCLUTION

This study demonstrates that critical thinking skills have a significant influence on pre-service teachers' ability to construct higher-order Indonesian language questions. Pre-service teachers with higher levels of critical thinking were able to formulate questions that were more analytical, evaluative, and creative. These findings confirm that critical thinking serves as a fundamental cognitive foundation in the development of higher-order questioning skills.

Overall, the study highlights the importance of cultivating critical thinking as a core professional competency for pre-service teachers. Strengthening this skill not only enhances the quality of assessment items but also supports more meaningful, contextual, and higher-order thinking–oriented Indonesian language learning.

Most participants were classified in the moderate category, indicating the need to reinforce critical and reflective thinking training within teacher education programs. The integration of inquiry-based learning models, metacognitive reflection, and technology such as AI-assisted scaffolding has been proven effective in various Scopus-indexed studies and is recommended for implementation.

Future research is encouraged to adopt multi-institutional or longitudinal designs, as well as mixed-method approaches that combine automated question analysis with expert evaluation, to gain a more comprehensive understanding of this relationship.

REFERENCE

Chan, C. C., Tsui, M. S., Chan, M. Y. C., & Hong, J. H. (2002). Applying the structure of the observed learning outcomes (SOLO) taxonomy on student's learning outcomes: An empirical study. *Assessment and Evaluation in Higher Education*, *27*(6), 511–527. https://doi.org/10.1080/0260293022000020282

Ennis, R. (2018). Critical thinking across the curriculum: A vision. *Springer*, *37*(1), 165–184. https://doi.org/10.1007/S11245-016-9401-4

Erdoğan, F., & Kalkan, S. (2024). Delving into the Critical Thinking Skills of Pre-Service Mathematics Teachers with Their Metacognitive Awareness. *Participatory Educational Research*, *11*(6), 280–299. https://doi.org/10.17275/per.24.90.11.6

Facione, P. A. (2011). Critical Thinking: What It Is and Why It Counts. In *Insight assessment* (Vol. 5, Issue 1). Insight Assessment. www.insightassessment.com

Fitriani, H., Asy'Ari, M., Zubaidah, S., & Mahanal, S. (2018). Critical Thinking Disposition of Prospective Science Teachers at IKIP Mataram, Indonesia. *Journal of Physics: Conference Series*, *1108*(1). https://doi.org/10.1088/1742-6596/1108/1/012091

Halpern, D. F., & Dunn, D. S. (2023). *Thought and Knowledge: An Introduction to Critical Thinking*. Taylor & Francis Group.

Huang, J., & Sang, G. (2023). Conceptualising critical thinking and its research in teacher education: A systematic review. *Teachers and Teaching: Theory and Practice*, *29*(6), 638–660. https://doi.org/10.1080/13540602.2023.2212364

Ilhan, M., & Gezer, M. (2017). A comparison of the reliability of the SOLO- and revised Bloom's taxonomy-based classifications in the analysis of the cognitive levels of assessment questions. *Pegem Egitim ve Ogretim Dergisi*, *7*(4), 637–662. https://doi.org/10.14527/pegegog.2017.023

Küçükaydın, M. A. (2024). Modeling the relationship between academic self-efficacy, metacognitive thinking skills, career plan, and academic motivation. *Quality and Quantity*, *58*(2), 1113–1130. https://doi.org/10.1007/s11135-023-01691-y

Lai, W. Y. W., Kim, P., & Lee, J. S. (2025). Designing and validating an AI-supported tool for enhancing critical inquiry in EFL education. *Research Methods in Applied Linguistics*, *4*(3). https://doi.org/10.1016/j.rmal.2025.100266

Lodovico Molina, I., Švábenský, V., Minematsu, T., Chen, L., Okubo, F., & Shimada, A. (2024). Comparison of Large Language Models for Generating Contextually Relevant Questions. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *15160 LNCS*, 137–143. https://doi.org/10.1007/978-3-031-72312-4_18

Lu, D., & Zheng, X. (2024). Teacher Critical Thinking Questioning in EFL Classes in China High Schools. *Educational Studies - AESA*, *60*(6), 659–676. https://doi.org/10.1080/00131946.2024.2437639

Luo, S., Zhang, W., Zheng, R., & Wang, A. (2025). The Influence of Procedural Scaffolding on pre-Service Teachers' Questioning Skills in Human-AI Collaborative Environment. *2025 7th International Conference on Computer Science and Technologies in Education, CSTE 2025*, 191–195. https://doi.org/10.1109/CSTE64638.2025.11091925

Metin Peten, D. (2022). Influence of the Argument-Driven Inquiry with Explicit-Reflective Nature of Scientific Inquiry Intervention on Pre-service Science Teachers' Understandings About the Nature of Scientific Inquiry. *International Journal of Science and Mathematics Education*, *20*(5), 921–941. https://doi.org/10.1007/s10763-021-10197-8

Mondal, H. (2025). Ethical engagement with artificial intelligence in medical education. *Advances in Physiology Education*, *49*(1), 163–165. https://doi.org/10.1152/ADVAN.00188.2024

Moorhouse, B. L., Ho, T. Y., Wu, C., & Wan, Y. (2025). Pre-service Language Teachers' Task-specific Large Language Model Prompting Practices. *RELC Journal*. https://doi.org/10.1177/00336882251313701

Orakcı, Ş., & Khalili, T. (2025). The impact of cognitive flexibility on prospective EFL teachers' critical thinking disposition: the mediating role of self-efficacy. *Cognitive Processing*, *26*(1), 59–73. https://doi.org/10.1007/s10339-024-01227-8

Santos Meneses, L. F. (2020). Critical thinking perspectives across contexts and curricula: Dominant,

neglected, and complementing dimensions. *Thinking Skills and Creativity*, *35*. https://doi.org/10.1016/j.tsc.2019.100610

Saputri, W., Corebima, A. D., Susilo, H., & Suwono, H. (2020). Qasee: A potential learning model to improve the critical thinking skills of pre-service teachers with different academic abilities. *European Journal of Educational Research*, *9*(2), 853–864. https://doi.org/10.12973/eu-jer.9.2.853

Scaria, N., Dharani Chenna, S., & Subramani, D. (2024). Automated Educational Question Generation at Different Bloom's Skill Levels Using Large Language Models: Strategies and Evaluation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *14830 LNAI*, 165–179. https://doi.org/10.1007/978-3-031-64299-9_12

Vuleta, K., & Martinović, A. (2019). Teacher questions in EFL classrooms. *Metodicki Ogledi*, *26*(1), 149–175.https://www.scopus.com/inward/record.uri?eid=2-s2.0-85070694352&partnerID=40&md5=9d1c114486ba593843110b477083086e

Wulandari, D., Liliasari, S., & Widhiyanti, T. (2021). The effect of argument-driven inquiry on chemistry reaction-rates to enhance pre-service chemistry teachers critical thinking skills. *Journal of Physics: Conference Series*, *1806*(1). https://doi.org/10.1088/1742-6596/1806/1/012204

Xiao, X., Li, Y., He, X., Fang, J., Yan, Z., & Xie, C. (2025). An assessment framework of higher-order thinking skills based on fine-tuned large language models. *Expert Systems with Applications*, *272*. https://doi.org/10.1016/j.eswa.2025.126531

Yıldız-Feyzioğlu, E., & Kıran, R. (2022). Investigating the Relationships between Self-efficacy for Argumentation and Critical Thinking Skills. *Journal of Science Teacher Education*, *33*(5), 555–577. https://doi.org/10.1080/1046560X.2021.1967568

Zhang, Z., Zhang, E., Liu, H., & Han, S. (2024). Examining the association between discussion strategies and learners' critical thinking in asynchronous online discussion. *Thinking Skills and Creativity*, *53*. https://doi.org/10.1016/j.tsc.2024.101588

ACKNOWLEDGMENT

The authors address their gratitude to the Directorate General of Teachers and Education Personnel, Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, through the Teacher Professional Education (PPG) program at the Universitas Islam Malang, for providing the opportunity for researchers to conduct research on the implementation of teacher professional education programs. The authors would also like to express their gratitude to the leadership of the Faculty of Teacher Training and Education, Universitas Islam Malang, for facilitating the implementation of this research, as well as to the lecturers and administrative staff of the Teacher Professional Education (PPG) Program for their assistance and cooperation during the data collection process. Special appreciation is extended to all pre-service teacher participants of the PPG Program at Universitas Islam Malang, who willingly served as research respondents and contributed valuable time, insights, and engagement throughout the study.