



## The Effectiveness of HOTS-Based Practice Questions in Enhancing Students' Critical Thinking in Arabic Learning

Khusnul Masrifah<sup>1\*</sup>, Mohammad Ahsanuddin<sup>2</sup>, Moh. Ainin<sup>3</sup>

<sup>1,2,3</sup> Universitas Negeri Malang, Indonesia

Correspondence: [khusnul.masrifah.2302318@students.um.ac.id](mailto:khusnul.masrifah.2302318@students.um.ac.id)

### ARTICLE INFO

#### **Article History:**

*Submitted/Received 14 Jul 2025*

*First Revised 14 Aug 2025*

*Accepted 20 Oct 2025*

*Publication Date 31 Oct 2025*

#### **Keywords:**

*Arabic Language Learning*

*Critical Thinking*

*Evaluation*

*HOTS-Based Practice Questions*

### ABSTRACT

Conventional assessment instruments in Arabic learning often emphasize low-level cognitive skills, leaving a gap in measuring students' higher-order and critical thinking abilities. Addressing this issue, the present research evaluates the effectiveness of Higher Order Thinking Skills (HOTS)-based practice questions in enhancing students' critical thinking within Arabic language learning. Using a quasi-experimental design with a non-equivalent control group, two classes were involved: one receiving HOTS-based questions and the other conventional questions. Data were collected through pretests and posttests and analyzed using the Independent Sample T-Test and Cohen's *d* to determine the significance and effect size. Results revealed that the experimental group achieved a higher mean posttest score (82.67) compared to the control group (74.10), with a significance value of 0.000 and an effect size of 1.52, indicating a strong influence. These findings confirm that HOTS-oriented assessment can significantly improve students' analytical and evaluative thinking skills. Beyond statistical significance, this study highlights the pedagogical importance of integrating HOTS-based evaluation as a transformative strategy to foster deeper reasoning, reflective understanding, and learner autonomy in Arabic education.

## 1. INTRODUCTION

The ability to think critically is one of the essential competencies in 21st-century education, enabling students to analyze, evaluate, and create knowledge effectively (Al-Smadi, 2023; Asmarani & Rahman, 2023). In Arabic language learning, critical thinking extends beyond grammatical mastery and vocabulary acquisition; it also encompasses interpreting cultural context, analyzing implicit meanings, and constructing arguments based on deep understanding (Mansur & Hakim, 2023; Mahmud & Al-Qudah, 2023; Kamalz, 2021). Critical thinking integrates analytical, evaluative, and creative dimensions that align closely with the core principles of Higher Order Thinking Skills (HOTS), emphasizing reasoning, innovation, and problem-solving (Affandy et al., 2024; Hu & Bi, 2024).

The revised Bloom's Taxonomy identifies three upper-level cognitive domains: analysis, evaluation, and creation as the foundation of HOTS (Anderson & Krathwohl, 2010a). Rather than focusing on question difficulty, HOTS challenges learners to engage in critical and creative reasoning when solving problems (S. Fauziah et al., 2020). Various studies have demonstrated that integrating HOTS into instruction improves students' metacognitive awareness, motivation, and problem-solving capacity across disciplines (Kania & Kusumah, 2025; Liu et al., 2024). In language education, HOTS-based tasks can nurture analytical reading, reflective writing, and interpretive comprehension skills essential for mastering Arabic as both a linguistic and cultural system.

Despite the global recognition of HOTS, its application in Arabic learning remains underdeveloped. Research has focused mainly on instrument design or content analysis without evaluating its empirical impact on students' thinking skills. For instance, Danni (2022) developed Arabic HOTS questions limited to item construction, while Mustofa et al. (2022) reported significant gains in analytical thinking but with instruments of limited reliability. Other studies, such as Nurbayan & Sanusi (2025) and Alshammari et al. (2024), primarily examined validity and motivation aspects rather than measuring actual learning outcomes. More recent initiatives have incorporated AI-assisted tools such as ChatGPT and augmented reality to generate HOTS items efficiently (Lu et al., 2025; Vebibina et al., 2025), yet empirical testing in Arabic learning contexts is still minimal.

Existing findings also highlight the persistent imbalance between LOTS (Lower-Order Thinking Skills) and HOTS elements in Arabic evaluation. Studies by Damaianti et al., (2020) show that Arabic exams in Indonesian schools contain only about 40–70% HOTS-level items, limiting opportunities for students to engage in higher reasoning and problem-solving. Moreover, while digital learning tools such as Quizizz and game-based learning models (Al-Smadi, 2023), have enhanced motivation and engagement, few studies have systematically examined whether HOTS-based practice questions directly strengthen students' critical thinking in Arabic language learning.

This study seeks to address these limitations by empirically testing the effectiveness of HOTS-based practice questions in enhancing students' critical thinking in Arabic learning. Unlike previous research that focused on the design or validation of HOTS instruments, this study applies a quasi-experimental approach to evaluate their pedagogical impact. The

research contributes to both theory and practice by (1) providing validated and reliable HOTS-based question instruments, (2) offering empirical evidence of their effect on students' critical thinking, and (3) presenting pedagogical implications for teachers implementing the Independent Curriculum that emphasizes reasoning, reflection, and learner autonomy.

## 2. METHODS

This study employed a quasi-experimental design using a non-equivalent control group approach to examine the effect of Higher Order Thinking Skills (HOTS)-based practice questions on students' critical thinking in Arabic learning. This design was chosen because it allows a valid comparison between groups that cannot be randomly assigned. The study was conducted at SMA Muhammadiyah 1 Ponorogo, East Java, Indonesia, during the second semester of the 2024/2025 academic year (March–May 2025). The population comprised all Grade X students at SMA Muhammadiyah 1 Ponorogo. Using purposive sampling, two classes with equivalent Arabic achievement scores from the previous semester were selected: one as the experimental group ( $n = 30$ ) and one as the control group ( $n = 30$ ). The experimental group received HOTS-based practice questions, while the control group was taught using conventional exercises. Participation was voluntary, with written consent obtained from students and school authorities. All data were treated confidentially and used solely for research purposes in accordance with institutional ethical standards. The study utilized a pretest and posttest procedure for both groups to measure changes in students' critical thinking skills after the intervention. The structure of the design can be summarized as follows:

Table 1. *Research design*

<b>Group</b>	<b>Pre test</b>	<b>Treatment</b>	<b>Post test</b>
<b>Experimental</b>	01	X <sub>1</sub> (HOTS-based practice questions)	02
<b>Control</b>	03	X <sub>2</sub> (Conventional practice questions)	04

Critical thinking skills were assessed through essay-based test items developed according to HOTS indicators, analyzing, evaluating, and creating, as adapted from Anderson and Krathwohl's (2010) taxonomy. Three Arabic education experts reviewed the instrument's content validity, and revisions were made to ensure alignment with the learning objectives. Reliability analysis using Cronbach's Alpha produced coefficients of 0.936 for the HOTS-based question set, 0.971 for the conventional question set, and 0.931 for the critical thinking posttest instrument, indicating excellent internal consistency ( $\alpha > 0.90$ ). Both experimental and control groups completed a pretest measuring baseline critical thinking skills. The experimental group then received a structured treatment using HOTS-based practice questions designed to promote higher-level reasoning. At the same time, the control group continued using standard exercises emphasizing comprehension and recall. After three instructional sessions, both groups completed a posttest under identical conditions. Data were

analyzed using IBM SPSS Statistics version 20. Prior to hypothesis testing, the Kolmogorov–Smirnov test and Levene’s test were used to verify normality and homogeneity assumptions. An Independent Sample T-Test was conducted to determine significant differences between the experimental and control groups, with a significance level set at  $\alpha = 0.05$ . Additionally, Cohen’s *d* was calculated to assess the magnitude of the treatment effect, where  $d > 0.80$  was interpreted as a large effect size.

### 3. RESULTS AND DISCUSSION

#### 3.1 Results

##### Normality Test Results

The improvement of students' critical thinking skills was analyzed based on the results of the pretest and posttest; this was done through normality testing of the gain score data. This normality test is the initial stage before the testing of the central hypothesis, which aims to ensure the fulfillment of the assumption of normal distribution as a prerequisite for the use of parametric analysis. The test was carried out using the Kolmogorov–Smirnov method using the IBM SPSS Statistics version 20 application. The decision-making criteria are based on significance values (Sig.), where the data is categorized as normally distributed when the value exceeds the threshold of 0.05. If this condition is met, the data is declared suitable for further analysis with the Independent Sample T-Test. The results of the normality test are presented in the following sections:

Table 2. Normality Test Results

Variabel		Kolmogorov-Smirnov Z	Sig.	Conclusion
Improvement of Critical Thinking Skills in Experimental and Control Classes	Pretest	0,844	0,475	Normal Distributed
	Posttest	1,035	0,234	Normal Distributed

As a result of the test, the presented normality test table shows that the null ( $H_0$ ) hypothesis cannot be rejected. The significance value for the experimental group was 0.475, while the control group recorded a value of 0.234. Both values exceeded the threshold of 0.05, which signifies that the data from both groups follow a normal distribution. Thus, the basic assumptions for the use of parametric analysis have been met.

##### Homogeneity Test Results

The homogeneity test was conducted to identify the similarities of variance between the two groups involved in the study. This analysis aims to ensure that the two groups have comparable data dispersion characteristics, thus allowing for valid parametric tests. The homogeneity test was carried out through Levene's Test procedure using IBM SPSS Statistics software version 20. The decision-making criteria are based on significance values, where the variance between groups can be homogeneous if the value exceeds 0.05. The results of the homogeneity test are presented in the following sections:

Table 3. Homogeneity Test Results

Variabel	Class		Levene statistics	Sig.	Conclusion
Improvement of Critical Thinking Skills	Control and Experiment Class	Pretest	0,844	0,698	Homogeneous Normal Distributed
		Posttest	1,035	0,157	

The homogeneity test results table indicated that for the experimental group or the control group, a significance value above 0.05 was produced. These findings reflect that the variance between the two groups is homogeneous, so the assumption of similarity of variance in parametric analysis has been met.

### Independent Sample T-Test and Effect Size Test Results

The prerequisite testing of the analysis is completed, and then the hypothesis is tested with parametric statistics. This is done because the experimental and control class data are usually homogeneously distributed.

The hypothesis formulation in this study is as follows:

$H_0$  (Hypothesis zero): There was no significant effect of the use of HOTS-based practice questions on students' critical thinking skills in learning Arabic at SMA 1 Muhammadiyah Ponorogo.

$H_a$  (Alternative hypothesis): There is a significant influence of the use of HOTS-based practice questions on students' critical thinking skills in learning Arabic at SMA 1 Muhammadiyah Ponorogo.

The following are the results of the analysis that has been carried out;

Table 4. Independent *Sample T-Test* Test Results of Critical Thinking Skills Improvement Data

		<b>Independent Samples Test</b>						
		<i>Levene's Test for Equality of Variances</i>			<i>t-test for Equality of Means</i>			
		F	<i>Itself.</i>	t	Df	<i>Sig.(2-tailed)</i>	<i>Mean Difference</i>	<i>Std. Error Difference</i>
Critical Thinking Skills	<i>Equal variances assumed</i>	1,266	0,266	-5,027	46	0,000	-,19458	0,03871
	<i>Equal variances not assumed.</i>			-5,027	43,988	0,000	-,19458	0,03871

Based on the results of the analysis using the Independent Sample T-Test, a significant result of p (sig. 2-tailed) of 0.000 was obtained, which is below the significance limit of 0.05. From these results, the null hypothesis (H0) was rejected and the alternative hypothesis (H1) was accepted. This indicates that the use of Higher Order Thinking Skills (HOTS)-based practice questions has a significant influence on improving students' critical thinking skills at SMA 1 Muhammadiyah Ponorogo.

The next step is to calculate the effect size after performing the sample's independent t-test. Effect size calculation is used to find out how much the quantitative value of an effect occurs after a treatment is given. Table 19 shows the results of the calculation of the impact measure of students' critical thinking skills.

Table 5. Effect Size Test Results for Critical Thinking Skills Improvement

Xt	Xc	Spooled	d (Effect Size)
0,48	0,28	0,131255	1,523755

Based on the results of the calculation using *the effect size* test, a d value of 1.523755 was obtained. According to the interpretation criteria, Cohen's value d is included in the high criteria.

### 3.2 Discussion

The statistical results indicate a significant improvement in students' critical thinking skills after the implementation of HOTS-based practice questions, with a p-value of 0.000 and a large effect size ( $d = 1.52$ ). These findings confirm that the intervention produced not only measurable gains in test scores but also meaningful development in higher-order reasoning processes. In the context of Arabic learning, this suggests that assessment instruments designed with HOTS indicators such as analysis, evaluation, and synthesis can stimulate students to move beyond surface comprehension toward deeper interpretive engagement with Arabic texts.

The improvement observed in the experimental group demonstrates that the integration of HOTS-oriented questions transforms the traditional function of assessment from a summative tool into a formative and cognitive-learning strategy. Instead of merely recalling grammatical rules or memorizing vocabulary, students were encouraged to construct arguments, interpret context, and apply logical reasoning in answering questions. This aligns with Brookhart (2010) and Anderson & Krathwohl (2010b) theoretical framework, which emphasizes that the upper levels of Bloom's Taxonomy analysis, evaluation, and creation are essential for fostering advanced academic reasoning and transferable cognitive skills.

Compared with previous studies, this research provides a more comprehensive empirical perspective. For instance, Amin et al., (2023) confirmed the positive influence of HOTS-based questions on students' analytical and evaluative skills in Arabic learning.. However, their studies primarily emphasized motivational and practical aspects without quantifying effect size or comparing HOTS-based versus conventional questions under controlled conditions. The current study extends these findings by using a validated

instrument and robust statistical testing, thereby providing stronger evidence for the pedagogical effectiveness of HOTS in Arabic learning. Similarly, the results are consistent with Al-Smadi (2023), who found that HOTS-oriented evaluation enhances students' interpretive and reflective thinking. However, this study goes beyond descriptive confirmation by demonstrating that validated HOTS-based instruments significantly raise measurable indicators of critical thinking, not merely self-reported engagement.

The findings also address the concern raised by Damaianti et al., (2020) and Elmi et al. (2024), who reported that Arabic language assessments in Indonesian schools still contain less than 70% HOTS content. By empirically proving that well-designed HOTS questions lead to higher critical thinking gains, this study underscores the urgency of redesigning Arabic assessments to better reflect the goals of the Independent Curriculum. From a pedagogical standpoint, the use of HOTS-based evaluation promotes metacognition, self-regulation, and reflective learning competencies that are central to 21st-century education (Affandy et al., 2024; Yuniarti et al., 2024). Furthermore, the positive results support the claim that integrating HOTS-based instruction can reduce the dominance of rote memorization and cultivate interpretive literacy among Arabic learners (Mahmud & Al-Qudah, 2023; Mansur & Hakim, 2023).

Beyond cognitive development, the use of HOTS questions also fosters learner motivation and autonomy. Students engaged with contextual, problem-solving-based items demonstrated higher persistence and curiosity, aligning with findings from Lu et al. (2025) and Wu et al. (2025), who reported that HOTS-oriented learning supported by technology increases engagement and critical awareness. In this regard, HOTS-based questions not only measure knowledge but also serve as scaffolds for independent reasoning, encouraging learners to construct meaning rather than reproduce information.

In terms of theoretical contribution, this study strengthens the pedagogical link between HOTS assessment and critical thinking development in Arabic learning a field still underexplored compared to science or mathematics education. The validated HOTS-based question model developed here can serve as a practical framework for teachers seeking to align Arabic instruction with higher cognitive objectives. Moreover, the strong effect size obtained (1.52) indicates that HOTS integration can significantly enhance both the depth and quality of student reasoning, supporting the view that critical thinking is teachable and measurable within language learning contexts.

Finally, the novelty of this research lies in combining empirical rigor with contextual relevance. Unlike previous works limited to instrument design or qualitative observation, this study quantitatively verifies the impact of HOTS-based questions through experimental testing and validated measures. The results contribute to Arabic education by providing a replicable model for evidence-based assessment reform. Pedagogically, the study emphasizes that designing Arabic evaluations with HOTS elements not only raises learning outcomes but also prepares students to think critically, reason reflectively, and engage with linguistic and cultural complexities key competencies for global and independent learners.

#### 4. CONCLUSION

This study demonstrates that the use of HOTS-based practice questions can effectively foster students' critical thinking in Arabic language learning. By engaging learners in analytical, evaluative, and creative tasks, this approach transforms assessment from a mere measurement tool into a medium for cognitive development and reflective reasoning. The findings highlight that HOTS-oriented evaluation not only strengthens learners' ability to analyze and synthesize information but also cultivates deeper understanding and independent thinking skills essential for 21st-century education. The novelty of this research lies in its empirical validation of HOTS-based assessment within the context of Arabic language education, an area that remains underexplored. Through a quasi-experimental design supported by reliable instruments, this study contributes substantive evidence to the growing body of research on pedagogical innovation in Arabic learning. The results underscore the need for teachers and curriculum designers to adopt HOTS-based strategies that promote reasoning, interpretation, and learner autonomy in classroom assessment practices. In practical terms, educators are encouraged to integrate HOTS-oriented evaluations in daily instruction to enhance student engagement and higher-order cognition. Future studies may extend this work by exploring long-term implementation of HOTS-based learning, utilizing digital and AI-assisted tools, and examining their broader effects on creativity, motivation, and metacognitive awareness in Arabic and other language education contexts.

#### 5. REFERENCES

- Al-Smadi, M. (2023). Promoting critical thinking through task-based learning in Arabic as a foreign language classrooms. *Arab World English Journal*, 14(4), 356–372. <https://doi.org/10.24093/awej/vol14no4.24>
- Alshammari, A., Shaalan, D., & Al-onazi, B. (2024). Instruments measuring motivation to learn Arabic as a second language: evidence of validity and reliability. *Humanities and Social Sciences Communications*, 11(1), 1-11. <https://doi.org/10.1057/s41599-024-03381-y>
- Amin, N. F., Nasruni, N., Zainuddin, N., & Mutmainnah, A. (2023). Higher, Medium and Lower Order Thinking Skills in the Book al-'Arabiyah Baina Yadaik. *Arabiyatuna: Jurnal Bahasa Arab*, 7(2), 795–816. <https://doi.org/10.29240/jba.v7i2.8414>
- Anderson, L. W., & Krathwohl, D. R. (2010a). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Anderson, L. W., & Krathwohl, D. R. (2010b). *Kerangka Landasan untuk Pembelajaran, Pengajaran, dan Asesmen: Revisi Taksonomi Bloom*. Yogyakarta: Pustaka Pelajar.
- Asmarani, S., & Rahman, A. (2023). Higher-order thinking skills in Arabic language learning: A systematic review of instructional practices in Indonesian schools. *Arabiyat: Journal of Arabic Education*, 10(2), 145–162. <https://doi.org/10.15408/a.v10i2.29781>
- Brookhart, S. M. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom*. Alexandria, VA: ASCD.
- D'Northwood, G., & Rattray, J. (2025). What is this thing called critical thinking? Perspectives

- from business school academics. *Innovations in Education and Teaching International*, 62(4), 1363-1376. <https://doi.org/10.1080/14703297.2024.2422310>
- Damaianti, V. S., Abidin, Y., & Rahma, R. (2020). Higher order thinking skills-based reading literacy assessment instrument: An Indonesian context. *Indonesian Journal of Applied Linguistics*, 10(2), 513–525. <https://doi.org/10.17509/ijal.v10i2.28600>
- Danni, R. (2022). Pengembangan Instrumen Penilaian HOTS pada Mata Pelajaran Bahasa Arab di MA Provinsi Kepulauan Bangka Belitung. *Afeksi: Jurnal Ilmiah Pendidikan Bahasa Dan Sastra Indonesia*, 9(1), 52–62. <https://doi.org/10.37149/afeksi.v9i1.308>
- Dibek, M. I., Kursad, M. S., & Erdogan, T. (2025). Influence of artificial intelligence tools on higher order thinking skills: a meta-analysis. *Interactive Learning Environments*, 33(3), 2216–2238. <https://doi.org/10.1080/10494820.2024.2402028>
- Elmi, B. A. O. N., Baroroh, R. U., Labibah, F. S., & Azzahra, L. Q. (2024). Innovation of HOTS-Based Test Instruments for Receptive Skills in Arabic Textbooks based on Kurikulum Merdeka. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 16(2), 577-597. <https://doi.org/10.24042/albayan.v16i2.24697>
- Fauziah, S., Kurniawati, R. D., & Wulandari, A. (2020). Peningkatan Keterampilan Berpikir Kritis melalui Penerapan Soal HOTS pada Pembelajaran IPA. *Jurnal Pendidikan Sains Indonesia*, 8(2), 251–260. <https://doi.org/10.24815/jpsi.v8i2.17463>
- Hu, X., & Bi, H. (2025). Exploring and validating the componential model of students' scientific critical thinking in science education. *Thinking Skills and Creativity*, 55, 1-8. <https://doi.org/10.1016/j.tsc.2024.101695>
- Kania, N., & Kusumah, Y. S. (2025). The Measurement of Higher-Order Thinking Skills: A Systematic Literature Review. *Malaysian Journal of Learning and Instruction*, 22(1), 97-116. <https://doi.org/10.32890/mjli2025.22.1.6>
- 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), 1–28. <https://doi.org/10.1186/s41239-024-00447-4>
- Liu, J., Liu, Z., Wang, C., Xu, Y., Chen, J., & Cheng, Y. (2024). K-12 students' higher-order thinking skills: Conceptualization, components, and evaluation indicators. *Thinking Skills and Creativity*, 52, 1-7. <https://doi.org/10.1016/j.tsc.2024.101551>
- Lu, Z., Chiu, M. M., Wang, S., Mao, W., & Lei, H. (2025). Effects of augmented reality on students' higher-order thinking: A meta-analysis. *Journal of Educational Computing Research*, 63(3), 728-747. <https://doi.org/10.1177/07356331241309644>
- Mahmud, E., & Al-Qudah, M. (2023). Fostering critical literacy in Arabic reading through inquiry-based instruction. *Journal of Language and Linguistic Studies*, 19(1), 142–158. <https://doi.org/10.52462/jlls.254>
- Mansur, M., & Hakim, L. (2023). Critical literacy approach in Arabic learning to support the Merdeka Curriculum. *Arabiyatuna: Jurnal Bahasa Arab*, 7(2), 121–138. <https://doi.org/10.29240/jba.v7i2.4786>

- Mustofa, S., Desrani, A., & Ritonga, A. W. (2022). HOTS in Arabic learning: a study of the implementation of HOTS on students' critical thinking ability. *Al-Ta'rib: Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya*, 10(2), 133-144. <https://doi.org/10.23971/altarib.v10i2.4088>
- Nurbayan, Y., & Sanusi, A. (2025). Integrating character values and EdTech media: A study of Arabic teachers' pedagogical competence in Bandung Raya. *International Journal of Learning, Teaching and Educational Research*, 24(8), 248-271. <https://doi.org/10.26803/ijlter.24.8.11>
- Vebibina, A., Rusijono, R., Mariono, A., Muafa, A., Kurniawan, D., Dewantoro, R. S., ... & Suseno, J. (2025). 11 Key Strategies in Using AI Chat GPT to Develop HOTS-Based Entrepreneurship Questionnaires. *Qubahan Academic Journal*, 5(1), 33-62. <https://doi.org/10.48161/qaj.v5n1a1134>
- Yuniarti, N., Rahmawati, Y., Anwar, M., Al Hakim, V. G., Hidayat, H., Hariyanto, D., Husna, A. F., & Wang, J. H. (2024). Augmented reality-based higher order thinking skills learning media: Enhancing learning performance through self-regulated learning, digital literacy, and critical thinking skills in vocational teacher education. *European Journal of Education*, 59(4),1-12. <https://doi.org/10.1111/ejed.12725>