



Higher Order Thinking Skills Development in Arabic Language Learning: A Systematic Literature Review

Masriah¹, Sopwan Mulyawan², Nursahid³

^{1,2,3} UIN Siber Syekh Nurjati Cirebon, Indonesia

¹ masriah@uinsssc.ac.id, ² sopmulyawan@gmail.com, ³ nshaa2205@gmail.com

Abstract

Higher order thinking skills (HOTS) have become a central policy target in Arabic language education in Indonesia, yet evidence on how HOTS is actually developed and assessed in classrooms remains scattered across small-scale studies. This systematic literature review synthesizes empirical studies on HOTS development in Arabic language learning at the secondary education level (MTs/MA/SMA), published between 2016 and 2026. Following PRISMA 2020 guidance with a Population-Concept-Context framework, searches through AI-assisted academic search tools and web-based Indonesian repositories identified 19 records, of which seven empirical studies met the inclusion criteria after title-abstract and full-text screening. A narrative thematic synthesis revealed three recurring patterns: a persistent gap between HOTS labels and actual assessment levels, media- and strategy-based efforts to cultivate HOTS, and structural challenges facing teachers and students. The review highlights that policy intent has outpaced classroom implementation and identifies concrete directions for future research and teacher training.

Keywords

Higher Order Thinking Skills; Arabic Language Learning; Systematic Literature Review; Secondary Education; PRISMA

مستخلص البحث

أصبحت مهارات التفكير العليا (HOTS) هدفاً سياسياً مركزياً في تعليم اللغة العربية في إندونيسيا، غير أن الأدلة حول كيفية تطويرها وتقييمها فعلياً في الفصول الدراسية لا تزال متناثرة بين دراسات صغيرة ومتفرقة. تهدف هذه المراجعة الأدبية المنهجية إلى تجميع الدراسات التجريبية حول تطوير مهارات التفكير العليا في تعلم اللغة العربية في المرحلة الثانوية، المنشورة بين عامي 2016 و2026. باتباع إرشادات PRISMA 2020 وإطار السكان-المفهوم-السياق، تم تحديد 19 دراسة عبر أدوات بحث أكاديمية مدعومة بالذكاء الاصطناعي ومستودعات إندونيسية على الإنترنت، وبعد الفرز اجتازت سبع دراسات تجريبية معايير الإدراج. كشف التحليل السردي الموضوعي عن ثلاثة أنماط متكررة، فجوة مستمرة بين تسمية الأسئلة بأنها من مستوى HOTS ومستواها الفعلي، وجهود قائمة على الوسائط والاستراتيجيات لتنمية هذه المهارات، وتحديات هيكلية يواجهها المعلمون والطلاب. تُظهر المراجعة أن نية السياسات سبقت التطبيق الصفي، وتحدد اتجاهات بحثية وتدريبية مستقبلية.

مهارات التفكير العليا؛ تعلم اللغة العربية؛ المراجعة الأدبية المنهجية؛ التعليم الثانوي؛ PRISMA

كلمات
أساسية

1. Background and Significance

Higher order thinking skills (HOTS), commonly operationalized through the analyze, evaluate, and create levels of the revised Bloom taxonomy, have been positioned as a core competency target across Indonesian basic education, including Arabic language instruction in madrasah. National curriculum instruments such as the Decree of the Minister of Religious Affairs (KMA) Number 183 of 2019 explicitly call for Arabic learning outcomes that go beyond memorization toward problem solving, critical thinking, and reasoning (Muradi et al., 2020). This policy direction reflects a broader concern that Indonesian students' higher order thinking performance has remained comparatively low on international benchmarks, motivating renewed attention to how HOTS is taught and assessed in specific subject areas, including Arabic as a foreign or second language.

2. Previous Research

A growing body of small-scale studies has examined HOTS in Arabic language education in Indonesia from different angles. Curriculum-document analyses have assessed whether official basic competencies accommodate HOTS categories (Muradi et al., 2020). Classroom-based studies have investigated specific media and strategies, such as crossword puzzles for reading comprehension (Ritonga et al., 2021a), as well as the challenges teachers faced implementing HOTS-based Arabic instruction during the Covid-19 pandemic (Ritonga et al., 2021b). Assessment-oriented studies have analyzed the quality and cognitive level of Arabic test items labeled as HOTS, both at the semester-examination level (Fauziah et al., 2020) and the thematic-question level (Kafi & Ilma, 2022), while more recent studies have examined teachers' strategies for designing HOTS-based questions under the Kurikulum Merdeka (Ashfia et al., 2024). A related but distinct strand of work has used Bloom's revised taxonomy to map official Arabic learning outcomes across educational levels without evaluating classroom implementation (Maulana, 2022).

3. Research Gap

Despite this accumulating literature, no study to date has systematically synthesized empirical findings on HOTS development specifically in Arabic language learning at the secondary education level (MTs/MA/SMA) using a transparent, replicable search and screening protocol. Existing reviews in this area are narrower in scope: Maulana (2022) synthesizes curriculum-outcome documents across the elementary and junior-secondary levels through a taxonomy lens without addressing classroom implementation, while other narrative overviews focus on technology trends or general digital-era opportunities rather than on how HOTS is concretely developed, assessed, and constrained in practice. The present review

addresses this gap by systematically searching, screening, and thematically synthesizing empirical studies focused on secondary-level Arabic learners, following PRISMA 2020 reporting guidance.

4. Research Objectives and Contribution

This study aims to (1) map the empirical literature on HOTS development in Arabic language learning at the secondary education level published between 2016 and 2026, and (2) identify recurring patterns across this literature regarding assessment practice, instructional strategy, and implementation challenges. The contribution of this review is threefold: it offers a PRISMA-guided synthesis specific to this population and concept, it makes explicit the methodological quality of the underlying studies through structured appraisal, and it identifies concrete, evidence-based directions for future primary research and teacher professional development.

5. Article Organization

The remainder of this article is organized as follows. The Method section describes the review protocol, search strategy, eligibility criteria, and quality appraisal tools. The Results and Discussion section reports the PRISMA flow, study characteristics, quality appraisal outcomes, and the three themes identified through narrative synthesis, discussed in relation to prior literature. The Conclusion section summarizes the main findings and outlines directions for further research.

طريقة البحث

This study is a systematic literature review (SLR) following PRISMA 2020 reporting guidance, using narrative thematic synthesis rather than statistical meta-analysis, because the included studies varied substantially in design (qualitative case study, document analysis, and quantitative item analysis) and did not report comparable effect sizes.

The review question was framed using a Population-Concept-Context (PCC) framework appropriate for a scoping-style synthesis: Population, secondary-level Arabic language learners (MTs/MA/SMA or equivalent); Concept, the development or improvement of higher order thinking skills in Arabic language learning, including instructional strategy, media, assessment, and curriculum; Context, formal Arabic language instruction, published between 2016 and 2026, with no restriction on article language.

Searches were conducted through two AI-assisted academic search tools (Consensus and a full-text semantic search corpus) and supplementary web searches targeting Indonesian repositories such as Garuda and SINTA, using English and Indonesian search terms. Manual export from named bibliographic databases was considered but not conducted, a decision made jointly with the

reviewing student after the automated search reached practical saturation for this review's scope; this is acknowledged as a limitation in the Conclusion section. A total of 19 unique records were identified after de-duplication.

Records were screened against inclusion criteria (empirical study on HOTS development among secondary-level Arabic learners, published 2016-2026, full text accessible) and exclusion criteria (wrong education level, wrong subject/concept focus, non-empirical or purely conceptual studies, or unverifiable publication venue). Nine records were excluded at this stage. Ten reports were sought for full-text assessment; three were set aside as related literature reviews rather than primary data sources and discussed narratively rather than pooled with the primary studies. Seven studies were included in the final synthesis.

Methodological quality was appraised using the CASP Qualitative Checklist for six qualitative studies and the JBI Checklist for Analytical Cross-Sectional/Descriptive Studies for one quantitative descriptive study, with any item unanswerable from the available report text marked "unable to assess" rather than assigned a default rating. Data extracted for each study included authorship, publication source, educational setting, study design, Arabic skill or assessment focus, sample, and key findings.

النتائج والمناقشة

Search and Selection Results

Figure 1 presents the PRISMA 2020 flow diagram. Of 19 records identified (11 via Consensus, 8 via web search of Indonesian sources), no duplicates were found across channels. Nineteen records were screened, and nine were excluded for reasons including educational level outside the target range, concept mismatch, non-empirical design, or unverifiable publication venue. Ten reports were assessed in full text; three were set aside as related literature reviews. Seven empirical studies were included in the final narrative synthesis.

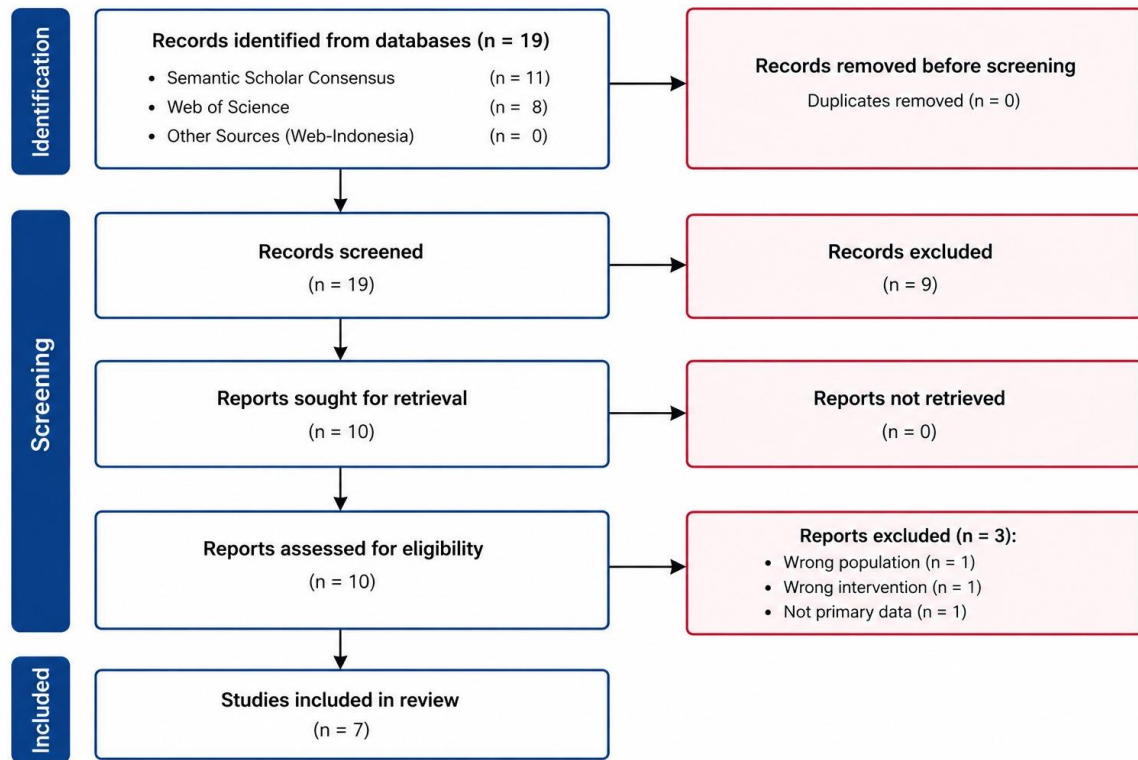


Figure 1. PRISMA 2020 flow diagram of the search and selection process.

Characteristics of Included Studies

Table 1 summarizes the seven included studies. Six were qualitative (document analysis or case study); one (Kafi & Ilma, 2022) applied quantitative item analysis. Settings ranged from a curriculum-policy level (Muradi et al., 2020) to specific MTs and MAN sites across Java and West Sumatra.

Study	Setting / Level	Design	HOTS Focus	Key Finding
Muradi et al. (2020)	Madrasah (curriculum document, KMA 183/2019)	Qualitative document analysis	Basic competency vs. HOTS taxonomy	Accommodates problem solving, critical thinking, reasoning; not yet creative thinking/decision making
Ritonga et al. (2021a)	SMA Islam Sabilillah Malang	Qualitative case study	Reading (qira'ah) media	Crossword puzzle media met HOTS indicators; student-centered, creative engagement
Ritonga et al. (2021b)	MAN 3 Padang Panjang	Qualitative case study	Teacher implementation challenges	Three barriers: student proficiency/access, teacher understanding, school policy
Fauziah et al. (2020)	MTs Al-Musyawah, Lembang (n=30)	Quantitative item analysis	Semester exam items	High validity/reliability, but all items classified as LOTS despite HOTS label
Kafi & Ilma (2022)	MAN 3 Jember (n=101)	Quantitative item analysis	Thematic exam items	31 of 35 items valid (89%); item quality did not fully reflect HOTS

Study	Setting / Level	Design	HOTS Focus	Key Finding
Ashfia et al. (2024)	MTs Pembangunan Jakarta	Qualitative case study	Teacher strategy for item design	Interactive media used; items leaned toward analyze level; time and adaptation challenges
Febriza et al. (2025)	MTs Ar Rohmah Petapahan	Qualitative content analysis	PTS essay items	8 of 10 essay items met HOTS criteria; 2 did not

Quality Appraisal

Appraisal using CASP and JBI checklists showed a consistent pattern across the seven studies: clarity of research aims, appropriateness of qualitative or descriptive methodology, and clarity of reported findings were generally well supported. In contrast, reporting on researcher reflexivity, ethical considerations, and detailed sampling strategy was largely absent from the published reports, a common reporting gap in this regional education literature rather than necessarily a flaw in the underlying research design. Reporting was comparatively strongest for Ritonga et al. (2021a, 2021b) and Fauziah et al. (2020), which described data collection and analysis procedures explicitly; reporting was comparatively thinner for Muradi et al. (2020) and the two MTs-based item-analysis studies. This appraisal informed how much interpretive weight each study was given in the synthesis below.

Theme 1: A Persistent Gap Between HOTS Labels and Actual Assessment Levels

Three studies converge on a striking pattern: instruments and competencies labeled or intended as HOTS-based frequently do not reach HOTS cognitive levels once analyzed against Bloom's revised taxonomy. Muradi et al. (2020) found that Arabic basic competencies in KMA 183/2019 accommodate problem solving, critical thinking, and reasoning, but not yet creative thinking or decision making. Fauziah et al. (2020) analyzed a full semester examination at one MTs and found that, despite being framed as HOTS-based, every item fell into the LOTS category (C1-C3) after taxonomic classification. Kafi and Ilma (2022) reported a related, though less stark, pattern in thematic test items at MAN 3 Jember, where item quality did not yet fully reflect higher order cognitive demand. Read together, these findings suggest that the gap lies less in curriculum intent and more in test-construction practice, indicating that teacher training in item writing, not only in HOTS conceptual understanding, is a priority.

Theme 2: Media- and Strategy-Based Efforts to Cultivate HOTS

Two studies documented concrete instructional efforts to develop HOTS beyond assessment redesign. Ritonga et al. (2021a) found that crossword puzzle media used in a reading comprehension (maharah qira'ah) lesson at one SMA met HOTS indicators, promoting student-centered, creative engagement with text rather than rote vocabulary recall. Ashfia et al. (2024) reported that Arabic teachers at one MTs used interactive media, including video and case-based tasks, to connect

Arabic content to students' real-life contexts, though item construction still leaned toward the analyze level rather than evaluate or create. Together, these studies indicate that pedagogical strategy and media choice offer a more immediately actionable lever for HOTS development than curriculum redesign alone, complementing the assessment-focused findings in Theme 1.

Theme 3: Structural Challenges Facing Teachers and Students

Two studies reported barriers to HOTS implementation at the classroom and institutional level. Ritonga et al. (2021b) identified three interacting challenges during pandemic-era online Arabic instruction: students' limited Arabic proficiency and internet access, teachers' incomplete conceptual understanding of HOTS, and school policies not yet aligned with HOTS standards. Ashfia et al. (2024) reported more localized constraints, limited instructional time and students' unfamiliarity with HOTS-style questions, even in an in-person, post-pandemic setting. These findings suggest that implementation barriers are not solely a pandemic-era artifact but persist structurally, spanning teacher capacity, student readiness, and institutional policy.

Relation to Related Literature Reviews

The three related reviews set aside during screening reinforce and contextualize these themes rather than duplicate them. Maulana (2022), synthesizing official learning-outcome documents through Bloom's revised taxonomy, found that Arabic learning outcomes at the elementary level (Fase B) remain at the LOTS level, while junior-secondary outcomes (Fase D) mix LOTS for listening/reading with HOTS-level creation targets for speaking and writing, a policy-level pattern consistent with the classroom-level gap documented in Theme 1. Two further reviews addressed technology integration and Kurikulum Merdeka policy context for HOTS-based Arabic instruction; both were library-research syntheses rather than primary empirical studies and are cited here as background rather than as data sources, consistent with this review's inclusion criteria.

الاستنتاج والخلاصة

This systematic literature review synthesized seven empirical studies on the development of higher order thinking skills in Arabic language learning at the secondary education level in Indonesia, published between 2016 and 2026. Three consistent patterns emerged: a persistent gap between the HOTS label attached to curricula and test instruments and the cognitive level these instruments actually reach; concrete, promising efforts by teachers to cultivate HOTS through media and instructional strategy; and structural challenges spanning teacher understanding, student readiness, and institutional policy that constrain implementation regardless of curriculum intent.

These findings carry a clear practical implication: policy statements mandating HOTS-based Arabic instruction, such as KMA 183/2019 and the Kurikulum Merdeka, are unlikely to translate into classroom practice without sustained teacher training specifically in HOTS-aligned item construction and pedagogy, rather than conceptual exposure to HOTS alone.

This review has clear limitations that should guide future work. First, the search retrieved only seven eligible empirical studies; a subsequent citation-chasing round identified at least five additional potentially eligible studies that were not incorporated in this review at the reviewing student's discretion, which limits the comprehensiveness of the synthesis and should be addressed in future updates. Second, no restriction was placed on article language, but in practice the review team could only assess studies in Indonesian, Arabic, and English with confidence; studies in other languages, if any exist, may not have been adequately captured. Third, several included reports provided limited methodological detail, particularly regarding sampling strategy, researcher reflexivity, and ethical procedure, which constrains confidence in some individual findings even where the overall thematic pattern across studies is consistent. Future research should pursue a more exhaustive search including manual database export, examine HOTS development in Arabic learning at the elementary and tertiary levels for comparison, and prioritize intervention studies that test specific HOTS-oriented teacher training programs rather than describing existing practice alone.

كلمة الشكر والتقدير

The author extends sincere appreciation and gratitude to the Arabic Language Education Department (PBA) at the State Islamic University (UIN) Siber Syekh Nurjati Cirebon (UIN SSC Cirebon) for the academic support, facilities, and opportunities provided, which enabled the successful completion of this research. Deep appreciation is also expressed to all lecturers in the PBA Department who offered direction, guidance, and valuable insights throughout the research process. Furthermore, the author thanks the PBA Department students who participated in this study by actively assisting with data collection and contributing to the editing and refinement of the manuscript. All the support, cooperation, and contributions provided played a vital role in the completion of this research.

المراجع

- Ashfia, A., Ridlo, U., & Raswan. (2024). Optimalisasi Higher Order Thinking Skill (HOTS) dalam Kurikulum Merdeka: Strategi dan konsep penyusunan soal bahasa Arab di MTs Pembangunan Jakarta. *Mauriduna: Journal of Islamic Studies*, 5(2), 248-260. <https://doi.org/10.37274/mauriduna.v5i2.1189>
- Fauziah, I. R. N., Syihabudin, S., & Sopian, A. (2020). Analisis kualitas tes bahasa Arab berbasis Higher Order Thinking Skill (HOTS). *LISANUNA: Jurnal Ilmu Bahasa Arab dan Pembelajarannya*, 10(1), 45-54. <https://doi.org/10.22373/ls.v10i1.7805>
- Febriza, S. I., Hikmah, & Masrun. (2025). Analysis of HOTS-based Arabic language PTS questions for class VII MTS Ar Rohmah Petapahan. *Al-Madrasah: Jurnal Pendidikan, Pembelajaran dan Kebudayaan*, 2(1), 14-27. <https://journal.syamilahpublishing.com/madrasah/article/view/276>
- Kafi, F. A., & Ilma, Z. (2022). Analisis soal tematik bahasa Arab berbasis HOTS pada kelas XII MAN 3 Jember. *Al-Fusha: Arabic Language Education Journal*, 4(1), 1-9. <https://doi.org/10.36835/alfusha.v4i1.729>
- Maulana, R. (2022). Analisis capaian pembelajaran bahasa Arab dengan Taksonomi Bloom Revisi. *Jurnal PTK dan Pendidikan*, 8(2). <https://doi.org/10.18592/ptk.v8i2.7621>
- Muradi, A., Mubarak, F., Darmawaty, R., & Hakim, A. R. (2020). Higher order thinking skills dalam kompetensi dasar bahasa Arab. *Arabi: Journal of Arabic Studies*, 5(2), 177-190. <https://doi.org/10.24865/ajas.v5i2.293>
- Ritonga, A. W., Ritonga, M., Septiana, V. W., & Mahmud. (2021a). Crossword puzzle as a learning media during the covid-19 pandemic: HOTS, MOTS or LOTS? *Journal of Physics: Conference Series*, 1933(1), 012126. <https://doi.org/10.1088/1742-6596/1933/1/012126>
- Ritonga, A. W., Wargadinata, W., Hasan, N., & Ahmad, B. M. B. (2021b). Teacher's challenges in implementing HOTS in learning Arabic during Covid-19 pandemic. *Izdiyar: Journal of Arabic Language Teaching, Linguistics, and Literature*, 4(1), 1-14. <https://doi.org/10.22219/jiz.v4i1.15606>