Journal of Arabic Language and Learning in Focus
e-ISSN: 2987-3347
Volume 2, Issue. 2, 2025, pp. 72-85
https://jurnal.isimupacitan.ac.id/index.php/alif.

Designing an Arabic Language Curriculum Based on Higher-Order Thinking Skills to Enhance the Critical Thinking Abilities of Madrasah Tsanawiyah Students

Hawinda Widya Fatma*[1], Aris Dianto[2], Hanif Adi[3]

[1,2]*Muhammadiyah University of Ponorogo, Indonesia.*
[3] *KarabükÜniversitesi, Türkiye.*
e-mail:
arisdianto2015@gmail.com
hawindafatma@gmail.com
hanifadipratama696@gmail.com

Submitted: 11-08-2026 Revised : 23-10-2026 Accepted: 30-10-2026

ABSTRACT. This study examines the design and implementation of a Higher-Order Thinking Skills (HOTS)-oriented Arabic language curriculum at MTsN 1 Pacitan, Indonesia. A qualitative case study approach was employed, using interviews, classroom observations, and document analysis involving teachers, curriculum stakeholders, and students. Analysis focused on curriculum planning, instructional interaction, learner responses, and assessment. The reported findings indicate efforts to connect learning objectives with contextual materials, text interpretation, discussion, analytical questions, and opportunities for students to justify their answers. Teachers and students described classroom practices that encouraged interpretation, comparison, evaluation, and reasoned responses. Differences in vocabulary knowledge, prior Arabic proficiency, and confidence constrained participation and highlighted the need for linguistic scaffolding. The findings are interpreted as evidence of opportunities to exercise critical thinking and of perceived changes in response patterns, rather than as proof of a causal increase in ability. The study argues for alignment among objectives, tasks, support, and assessment criteria that distinguish linguistic accuracy from the quality of reasoning. Its contribution is a situated account of curriculum implementation; incomplete reporting of sample and fieldwork details, together with the absence of longitudinal or comparative assessment, limits conclusions about effectiveness and transfer.

Keywords: *Arabic Language Curriculum; Higher-Order Thinking Skills; Critical Thinking; Madrasah Tsanawiyah.*

https://doi.org/10.52640/alif.304
How to Cite Fatma., H., W., et al. (2025). Designing an Arabic Language Curriculum Based on Higher-Order Thinking Skills to Enhance the Critical Thinking Abilities of Madrasah Tsanawiyah Students. *Journal of Arabic Language and Learning in Focus,* Volume 2(2), 72-85.

INTRODUCTION

An Arabic curriculum oriented toward higher-order thinking should make students' reasoning visible through what they do with language. Recognizing vocabulary and reproducing grammatical patterns remain necessary, but these accomplishments alone do not demonstrate that learners can compare interpretations, judge evidence, or defend a conclusion. Research on critical thinking in foreign-language education identifies explicit instruction, purposeful interaction, and assessment as recurring concerns, while also reporting considerable variation in instructional quality and research design (El Soufi & See, 2019; Yin et al., 2023). The curricular

problem is therefore how to connect language foundations with opportunities to reason, rather than how to replace foundational knowledge with increasingly difficult questions.

Higher-order thinking skills (HOTS) and critical thinking overlap but are not interchangeable. Analysis, evaluation, and creation describe demanding cognitive activity; critical thinking additionally requires attention to the quality of reasons and the adequacy of evidence. A difficult vocabulary exercise can remain a recall task, whereas comparing two accessible interpretations and defending one with textual evidence can require analysis. Language-specific work on critical-thinking taxonomies supports examining the actual reasoning demanded by a task rather than relying only on its instructional verb (Li & Liu, 2021). Reviews of textbook evaluation likewise caution that cognitive categories need explicit and consistently applied criteria (Soe, 2024).

Language teaching offers a productive setting for this distinction because learners must interpret meaning, relate statements to contexts, and decide whether an explanation is defensible. However, much of the available evidence concerns English as a foreign or second language rather than Arabic. Research on English classrooms identifies opportunities for critical discussion alongside constraints involving proficiency, classroom expectations, and teacher preparation (Liang & Fung, 2021; Yuan et al., 2022). These findings provide comparative pedagogical insights for Arabic education; they do not establish that the same outcomes will occur in an Indonesian madrasah. Arabic-specific curricular interpretation must attend to students' available vocabulary, their familiarity with written forms, and the communicative purposes of classroom texts.

Indonesian studies help specify how these issues appear in language education. Research on secondary-school language teachers found differences between their understanding of HOTS and their capacity to formulate higher-order questions in practice (Gozali et al., 2021). A study of junior secondary learners examined analogical reasoning through narrative texts, illustrating how text-based comparison can be organized as a learning activity (Suharno&Setyarini, 2021). Work on HOTS-oriented reading assessment also emphasizes the need to operationalize the construct in instruments appropriate to the Indonesian context (Damaianti et al., 2020). Taken together, these studies motivate an investigation of the relationship between intended curriculum, classroom tasks, and the evidence used to judge students' reasoning.

Curricular alignment matters because a higher-order objective can be weakened by a task that only requests a copied answer or an assessment that rewards linguistic accuracy alone. Teachers' understandings of critical thinking influence their selection of prompts, examples, and acceptable responses. Studies of both preservice teachers and classroom practitioners identify conceptual and contextual barriers to translating critical-thinking aspirations into teaching (Li, 2023; Yuan & Stapleton, 2020). Research from teachers' perspectives further supports attending to implementation rather than treating an announced curriculum as evidence of an enacted one (Cáceres et al., 2020). In Arabic learning, alignment therefore needs to be examined at the level of specific tasks and interactions.

Several instructional approaches offer useful comparisons. Indonesian studies have explored project-based critical reading, cooperative reading and composition, and structured discussion incorporating Socratic questioning (Lintangsari et al., 2022; Sari &Prasetyo, 2021; Susilo et al., 2021). These approaches differ in design and setting, but each draws attention to activities through which students can make and explain decisions. Their relevance to the present study lies in the proposed pedagogical processes, not in assuming that a particular teaching label guarantees critical thinking. Discussion, for example, can remain reproductive if students merely exchange memorized answers.

The same caution applies to technological approaches. Reviews of technology-supported language learning and ICT-assisted critical-thinking instruction identify opportunities as well as implementation challenges (Lu & Xie, 2024; Shadiev& Wang, 2022). An Indonesian junior-

secondary study of a mobile application is particularly relevant as an age-adjacent example, but its intervention and assessment differ from the present Arabic case (Agustina et al., 2022). Technology may support access to texts, feedback, or peer interaction, yet its presence cannot substitute for an explanation of how a task requires reasoning. The present study consequently treats technology as an optional instructional resource rather than a defining component of HOTS.

The contribution of this study is a contextual account of how curriculum planning, classroom interaction, and assessment relate within one madrasah. It does not claim that these relationships have never been studied or that the selected literature constitutes an exhaustive review. At MTsN 1 Pacitan, the central questions are how HOTS intentions are expressed in planning, how teachers and students enact them during Arabic learning, and what the available classroom evidence can reasonably indicate about critical thinking. A qualitative case study is appropriate to understanding these interrelated processes in their setting (Rashid et al., 2019).

Accordingly, this study examines the design and implementation of a HOTS-oriented Arabic curriculum at MTsN 1 Pacitan. Its analytic focus is the consistency among learning objectives, materials, teacher questioning, learner responses, and assessment. The study also considers how differences in Arabic proficiency affect students' opportunities to demonstrate reasoning. Its intended contribution is an evidence-conscious account of curriculum implementation and practical implications for teaching, rather than an estimate of the curriculum's causal effect on achievement.

METHOD

This study employed a qualitative case study approach to examine the design and implementation of a Higher-Order Thinking Skills (HOTS)-based Arabic language curriculum and its relationship to opportunities for students' critical thinking at MTsN 1 Pacitan, East Java, Indonesia. The research participants were selected purposively and consisted of the head of the madrasah, curriculum coordinator, Arabic language teachers, and students involved in Arabic language learning. Data were collected through in-depth interviews, classroom observations, and document analysis. Interviews explored curriculum planning, teachers' understanding and implementation of HOTS, instructional strategies, classroom challenges, and assessment practices. Classroom observations focused on teacher–student interactions, learning activities, questioning techniques, students' participation, and activities involving analysis, evaluation, problem solving, and conclusion drawing. Documents analyzed included curriculum documents, teaching modules, lesson plans, Arabic learning materials, worksheets, assessment instruments, and students' learning products. The design was used to examine implementation within its institutional context rather than to estimate an intervention effect (Rashid et al., 2019).

The reported analysis followed an interactive process comprising data condensation, data display, and conclusion drawing and verification. Interview transcripts, observation notes, and documents were coded and categorized according to the main themes of curriculum design, HOTS integration, critical thinking, instructional strategies, learning activities, and assessment. Data were subsequently presented through thematic descriptions and matrices to identify patterns between the planned curriculum and its classroom implementation. The credibility of the findings was established through source and technique triangulation, by comparing information from different participants and examining the consistency of interview findings with classroom observations and relevant documents. The analysis focused on identifying how HOTS principles were incorporated into curriculum objectives, learning processes, classroom interaction, and assessment, as well as how these practices created opportunities for students to develop analysis, evaluation, reasoning, and conclusion-drawing abilities. The comparison of interview, observation, and document evidence is consistent with using multiple sources to interpret a qualitative phenomenon (Busetto et al., 2020).

The methodological account does not specify participant numbers, the number or duration of observations, a precise fieldwork period, or an explicit procedure for determining data sufficiency. Their absence limits the assessment of data adequacy and transferability. The scope of the reported evidence is considered when interpreting the findings. Methodological research indicates that saturation and sample adequacy require justification in relation to the study's purpose and data characteristics (Hennink& Kaiser, 2022).

RESULTS AND DISCUSSION

Results

Curriculum Planning for HOTS-Oriented Arabic Learning

The first finding concerns the integration of HOTS into the planning of Arabic language learning. The curriculum planning process at MTsN 1 Pacitan refers to the applicable madrasah curriculum while adapting learning objectives and classroom activities to students' cognitive and linguistic development. The Arabic language teachers attempt to formulate learning activities that move beyond vocabulary memorization and grammatical recognition toward interpretation, analysis, evaluation, and production. This orientation can be identified in the teaching modules and learning plans, particularly through learning objectives that require students to understand information, identify relationships among ideas, interpret Arabic texts, and communicate their understanding.

"In planning the Arabic language curriculum, we guide teachers to ensure that learning objectives are not solely oriented towards content mastery but also focus on students' abilities to understand, analyse, evaluate, and apply the information they acquire. Therefore, HOTS components need to be incorporated from the initial stage of lesson planning, encompassing learning objectives, activities, and assessment." (Curriculum Coordinator, interview, 2025).

This finding was further reinforced by the Arabic language teacher, who explained that learning activities were designed to provide students with opportunities to reflect on the meaning and practical application of Arabic language materials. The teacher stated:

"When planning Arabic language lessons, I do not simply determine the materials and vocabulary that students need to master; I also consider how they can understand, analyse, and apply the materials in different situations. Thus, after developing a basic understanding of the subject matter, students are encouraged to engage in deeper thinking about what they have learned."(Arabic Language Teacher, interview, 2025).

Document analysis further indicated that HOTS-oriented elements were reflected in learning objectives, classroom tasks, and questions requiring students to provide explanations rather than merely reproduce information. The learning materials included activities in which students were asked to identify information from Arabic texts, compare linguistic or contextual information, explain the meaning of expressions, and formulate responses based on particular situations. These activities indicate an effort to align curriculum planning with higher-order cognitive processes.

Implementation of HOTS in Arabic Language Classrooms

The second finding concerns the implementation of the curriculum during classroom instruction. Classroom observations showed that Arabic learning involved a combination of teacher explanation, questioning, individual activities, group discussion, text interpretation, and presentation. The teacher did not rely exclusively on direct explanation but provided students with opportunities to respond to questions and discuss the meaning of Arabic expressions and texts. The learning process was particularly oriented toward encouraging students to explain the reasoning behind their answers.

One of the observed activities involved students working with Arabic texts and identifying information contained within them before explaining the relationship between the information and the learning topic. In several activities, the teacher used questions that required students to explain why an answer was considered appropriate rather than simply asking them to provide the correct answer. This classroom practice created opportunities for students to move from comprehension toward analysis and evaluation.

The Arabic language teacher explained that students initially tended to wait for the teacher's explanations before responding to questions. However, through repeated questioning and classroom discussions, students were gradually encouraged to articulate their own reasoning. The teacher stated: "At the beginning of the lessons, some students tended to wait for the teacher's explanations and immediately look for answers that they considered correct. Therefore, I tried to ask follow-up questions, such as why they had chosen a particular answer or what reasons supported their responses. Through this approach, students gradually became accustomed to explaining their thought processes rather than simply providing answers." (Arabic Language Teacher, interview, 2025).

The findings also indicate that questioning was one of the most frequently employed strategies for stimulating Higher-Order Thinking Skills (HOTS). The teacher posed questions concerning the meaning of texts, differences between expressions, the appropriate use of language in particular situations, and the reasons behind students' answer choices. Rather than immediately correcting incorrect responses, the teacher sometimes posed additional questions to encourage students to reconsider and evaluate their answers.

A student participant confirmed this classroom experience, stating:

"The teacher often asks not only what the answer is but also why we choose that particular answer. Sometimes, we are asked to compare two answers or explain the meaning of a sentence according to its context. Therefore, we have to think carefully before responding." (Student 1, interview, 2025).

These findings suggest that classroom interaction provided opportunities to exercise HOTS. Students were encouraged to interpret information and construct responses, while the teacher used follow-up questions to elicit and reconsider their reasoning.

Reported manifestations of students' critical thinking

The third finding relates to the emergence of critical thinking abilities during Arabic language learning. Based on classroom observations and interviews, students demonstrated several activities associated with critical thinking, particularly interpretation, analysis, evaluation, reasoning, and conclusion drawing. These abilities appeared when students were asked to interpret Arabic texts, distinguish relevant information, compare possible meanings, assess the appropriateness of an answer, and explain the reasons underlying their responses.

Interpretation was evident when students were required to understand the meaning of Arabic vocabulary, sentences, and short texts according to their context. Analysis appeared when students identified relationships among pieces of information or examined differences between Arabic expressions. Evaluation was demonstrated when students assessed whether a particular response was appropriate according to the context of a question or dialogue. Meanwhile, conclusion drawing appeared when students summarized information from Arabic texts or formulated responses based on the information they had analyzed.

The Arabic language teacher explained that the development of students' critical thinking skills could be observed through changes in their response patterns during classroom activities. According to the teacher: "The changes become evident when students begin to provide reasons to support their answers. Previously, they tended to respond based primarily on memorisation.

However, after being given questions that require analytical thinking, they have started to examine the context, identify relevant information, and explain why they choose particular answers. Although not all students have fully developed this ability, their progress in thinking skills has gradually become apparent." (Arabic Language Teacher, interview, 2025).

A student also explained that Arabic language learning activities required them to think before answering rather than simply memorise the learning materials. The student stated: "In our current lessons, we do not simply memorise vocabulary. After reading a text, we are usually asked to identify important information, explain its meaning, or provide answers supported by reasons. Therefore, before answering, we need to understand the text first." (Student 2, interview, 2025).

The findings therefore indicate that HOTS-oriented Arabic learning provides students with opportunities to engage in cognitive processes beyond remembering. The accounts describe students being exposed to learning situations in which answers had to be supported by explanations, contextual understanding, or logical reasoning. However, the development of critical thinking was not uniform among all students. Differences in vocabulary mastery, confidence, prior Arabic knowledge, and willingness to participate influenced students' ability to respond to higher-order questions.

Assessment of HOTS and Critical Thinking

The fourth finding concerns assessment. Assessment at MTsN 1 Pacitan was conducted through both learning-process activities and learning outcomes. The assessment included questions, assignments, classroom participation, discussions, and written tasks. The findings show that teachers attempted to incorporate higher-order questions into Arabic assessment, particularly through questions requiring students to interpret texts, apply language knowledge in contextual situations, and explain their answers.

The Arabic language teacher explained that the assessment process was designed to evaluate not only students' memorisation but also their ability to understand and apply knowledge in different contexts. The teacher stated: > "In designing assessments, I try not to rely solely on memorisation-based questions. Some questions require students to understand texts, identify information, apply vocabulary in different sentences, compare information, and provide reasons to support their answers. In this way, assessment does not merely measure whether students can recall the learning materials but also how effectively they can apply their knowledge." (Arabic Language Teacher, interview, 2025).

Document analysis indicated that assessment questions were not entirely limited to factual recall. Some tasks required students to apply vocabulary and grammatical knowledge in different contexts, interpret information from texts, and construct appropriate responses. This indicates an effort to create consistency between curriculum objectives, classroom activities, and assessment.

Nevertheless, the findings also reveal that the implementation of HOTS-based assessment still faces challenges. Teachers must balance higher-order questions with students' existing Arabic proficiency. Students who have limited vocabulary or grammatical knowledge may experience difficulties when confronted with complex questions, even when they possess adequate reasoning abilities. Therefore, the implementation of HOTS in Arabic learning requires gradual scaffolding so that students have sufficient linguistic resources to demonstrate their critical thinking.

Overall, the results demonstrate that the HOTS-based Arabic language curriculum at MTsN 1 Pacitan has been implemented through the alignment of curriculum planning, classroom interaction, learning activities, and assessment. The strongest evidence of implementation was found in the use of analytical questions, contextual tasks, text interpretation, classroom discussion, and opportunities for students to explain their reasoning. At the same time, the findings reveal that the reported implementation appears to depend on teachers' capacity to design appropriate learning experiences and provide scaffolding according to students' Arabic

proficiency. Thus, the curriculum functions not merely as a formal document but as a framework that guides teachers in creating Arabic learning experiences that connect linguistic practice with opportunities for critical thinking.

Discussion

Curriculum planning and the meaning of alignment

The planning evidence suggests that teachers at MTsN 1 Pacitan intend to connect Arabic knowledge with interpretation, comparison, and justified responses. The curriculum coordinator's account is important because it locates these intentions across objectives, activities, and assessment. Nevertheless, an intention to "understand" or "use" information does not itself establish HOTS. The analytical question is whether the task requires learners to organize relationships, evaluate alternatives against criteria, or construct an original response for a specified purpose. This distinction is consistent with language-specific critical-thinking taxonomies and research on teachers' interpretation of classroom thinking demands (Cáceres et al., 2020; Li & Liu, 2021).

The case therefore supports a qualified interpretation of alignment: relevant connections are described, but their consistency across all lessons is not established. A practical implication is to examine one learning objective together with its associated prompt, expected reasoning, and scoring criteria. For example, an objective about evaluating a response should be matched with competing responses and an opportunity to justify a choice, rather than with a vocabulary-definition question. This is a proposed curriculum-review procedure, not an additional finding from the fieldwork. Reviews of critical-thinking pedagogy and studies of Indonesian reading assessment support the importance of linking instructional intentions to observable performances (Damaianti et al., 2020; Yin et al., 2023).

Teacher questioning and opportunities to justify answers

Repeated follow-up questions were a prominent feature of the reported classroom interaction. Asking why a response fits a text can reveal whether a learner has inferred a relationship or merely repeated a teacher-provided answer. The teacher's decision to ask further questions before correcting an error also creates an opportunity for reconsideration. Indonesian research on HOTS questioning shows why this practice requires more than teachers' general agreement with the value of higher-order learning (Gozali et al., 2021). The local finding is strongest as evidence of an opportunity to reason; the quality of each student's reasoning still depends on the content of the response.

A useful progression is to ask students to identify a claim, locate supporting information, explain the link, and consider another plausible interpretation. These proposed prompts make the criteria for a defensible answer more explicit. Research on Socratic questioning within structured discussion offers a relevant comparison, although its online setting and research design differ from the present case (Lintangsari et al., 2022). Follow-up questioning should consequently be examined together with response time, linguistic support, and the teacher's handling of disagreement. Without these conditions, a nominally open question can still function as a search for one predetermined answer.

Discussion and collaborative reasoning

The observed discussions can be interpreted as spaces in which learners compare and explain meanings. Their potential contribution lies in the exchange of reasons rather than in group membership itself. Research on collaborative learning reports benefits for critical thinking, while work on cooperative reading and composition describes links between interaction and students' expression of an authorial position (Susilo et al., 2021; Warsah et al., 2021). These

comparisons help explain why peer discussion may be productive in Arabic learning, but they do not verify that every participant in the present case engaged at the same cognitive level.

Unequal participation remains an alternative explanation for apparently successful group work. A confident student may supply an answer that others repeat without understanding. The practical implication is to combine discussion with an individual explanation or a short response showing which textual evidence informed the group's decision. Project-based critical-reading research provides another example of organizing sustained learning around meaningful tasks (Sari &Prasetyo, 2021). For the present context, a modest, well-supported comparison task may be more informative than an elaborate project if it allows the teacher to examine each learner's reasoning.

Text interpretation and contextualized language use

The results indicate that students were asked to consider meanings in relation to a text or communicative situation. Contextualization can increase cognitive demand when learners must reconcile different pieces of information or choose between interpretations. However, identifying an explicitly stated detail is ordinarily evidence of comprehension, not automatically of analysis. Junior-secondary research on analogical reasoning is useful here because it makes comparison and the use of prior knowledge central to interpreting narrative content (Suharno&Setyarini, 2021). Work on language-specific critical-thinking categories similarly supports describing the reasoning operation rather than assigning a higher-order label to every text-based activity (Li & Liu, 2021).

A proposed Arabic task could present two possible replies to a short dialogue and ask students to choose the more appropriate reply, identify contextual clues, and explain why the alternative is less suitable. The linguistic content can remain familiar while the judgment becomes more demanding. Research on graphic organizers suggests that representing relationships can assist learners' analytical reading, while classroom studies emphasize the importance of supporting participation in a second language (Liang & Fung, 2021; Min et al., 2023). Such evidence supports trying accessible organizing tools; it does not establish that the same tool has already been used or evaluated at MTsN 1 Pacitan.

Instructional materials and the cognitive demand of tasks

The document analysis identifies questions that invite explanation and contextual responses. This is more informative than simply noting that a teaching module contains the term HOTS. Nonetheless, task classification requires examining what students actually need to do. A prompt beginning with "why" can still demand recall if its answer is already supplied, while a comparison of simple statements can require genuine analysis. A review of textbook evaluations found that higher-order demands were less commonly represented than lower-order ones and that methodological reporting in task classification was often limited (Soe, 2024). The present case should therefore avoid assigning cognitive levels solely from question stems.

An implication for materials development is to retain vocabulary and grammar support while adding opportunities to select evidence, compare interpretations, and justify a decision. For example, a worksheet can distinguish a preparatory comprehension question from a subsequent evaluative question instead of describing both as HOTS. Graphic-organizer research and the broader review of EFL writing pedagogy support examining how materials scaffold the target reasoning process (Min et al., 2023; Yin et al., 2023). This recommendation concerns the design of future materials; the available document descriptions do not establish the proportion of higher-order items in the existing curriculum.

Linguistic scaffolding and equitable participation

The reported differences in vocabulary, prior knowledge, and confidence suggest that weak Arabic output can have more than one cause. A learner may possess a defensible interpretation but lack the language needed to express it. Conversely, a grammatically accurate response may reproduce a memorized answer without independent judgment. Research on critical-thinking implementation in language classrooms and on teachers' conceptions underscores the interaction between pedagogical demands and the language-learning environment (Li, 2023; Liang & Fung, 2021). Interpreting performance therefore requires distinguishing limited access to the task from limited reasoning within the task.

Possible supports include a short vocabulary list, a worked example of using evidence, sentence starters for justification, and rehearsal before public speaking. These are proposals arising from the reported challenge, not additional observed procedures. The cognitive demand should remain visible: support can help a student articulate a reason without supplying the reason itself. Studies of flipped reading and listening illustrate alternative ways to provide preparation before demanding classroom activity (Etemadfar et al., 2020; Yulian, 2021). Their different designs do not demonstrate that flipping instruction is necessary; the transferable principle is providing adequate preparation and checking whether students can then reason with less assistance.

Metacognition and the interpretation of student development

The teacher's account of students beginning to explain their choices indicates a perceived change in classroom response patterns. This is educationally meaningful, but it should not be treated as a standardized measure of growth. More frequent explanations may reflect familiarity with the teacher's expectations, increased confidence, stronger vocabulary, or developing reasoning. Studies of metacognitive and critical-thinking practices in argumentative writing identify planning and monitoring as useful instructional targets (Murtadho, 2021; Pelenkahu et al., 2024). Their findings support examining learners' reasoning processes, while their different contexts limit direct comparison with Arabic learning at the junior-secondary level.

Future classroom assessment could ask learners to explain what changed between an initial answer and a revised answer, or which part of a text led them to reconsider a claim. Listening-journal research offers an example of structured reflection, but its reliance on self-assessment also shows why perceived progress should be distinguished from independently assessed performance (Purnamaningwulan, 2022). In the present study, interpretation, analysis, and justification are best reported as observed or reported manifestations of critical thinking. The evidence does not establish how much each learner improved or whether any improvement persisted beyond the observed lessons.

Assessment and feedback on the quality of reasoning

The assessment findings show an effort to include explanation and contextual application alongside language knowledge. This is compatible with evaluating reasoning, but the manuscript does not report a validated Arabic critical-thinking instrument or the detailed criteria used to score responses. Indonesian research on HOTS reading assessment demonstrates the importance of specifying and evaluating an instrument rather than assuming that open-ended tasks are sufficient (Damaianti et al., 2020). A proposed analytic rubric for Arabic learning could distinguish interpretation, relevance of evidence, coherence of justification, and consideration of alternatives from vocabulary and grammatical accuracy.

Separating these dimensions would make feedback more informative. A student could receive recognition for a reasonable inference while also receiving guidance on how to express it accurately in Arabic. A meta-analysis of educational feedback indicates that effects vary with the information conveyed, supporting feedback that identifies a gap and a next step rather than offering a score alone (Wisniewski et al., 2020). In the present setting, a comment such as asking

for the textual clue behind a conclusion would make the expected reasoning visible. This is an assessment recommendation, not a claim that a specific rubric or feedback intervention was tested in this study.

Peer review could complement teacher feedback if learners use clear criteria and have opportunities to revise. Evidence from a meta-analysis supports peer assessment as a formative practice for academic performance, but that outcome is broader than critical thinking and cannot be directly equated with it (Double et al., 2020). A manageable Arabic activity would ask a partner to identify the claim and evidence in a short response, then explain whether they match. Teacher moderation would still be needed to prevent incorrect language knowledge or peer agreement from becoming the sole basis for judging a response.

Teacher learning and sustainable curriculum implementation

The implementation described in the case depends on teachers' ability to interpret tasks, recognize evidence of reasoning, and respond to learner difficulty. These decisions cannot be reduced to inserting higher-order verbs into a lesson plan. Research on preservice teachers and a systematic review of EFL teachers' engagement with critical thinking identify gaps in conceptual understanding and support (Yuan & Stapleton, 2020; Yuan et al., 2022). Yan's discussion of teacher critical thinking and classroom engagement offers a conceptual rationale for examining this relationship, but it should not be presented as experimental evidence that teacher critical thinking caused the student responses reported here (Yan, 2021).

Collaborative lesson review is a plausible professional-development response. Teachers could examine a common text, predict students' reasoning and language difficulties, observe one another's questioning, and revise the task using examples of learner responses. A qualitative study of language teachers learning to teach critical thinking through lesson study provides a relevant model of professional learning (Zhang & He, 2023). Its small-scale contextual evidence supports the feasibility of reflective collaboration, not a guarantee of effectiveness across institutions. Madrasah support would need to include time for reviewing tasks and assessment examples, not only administrative completion of modules.

Digital resources as an optional extension

No evaluation of a digital intervention is reported in the present case. Technology should therefore be discussed as a possible extension rather than as an explanation for the findings. Reviews of technology-supported language learning and ICT-assisted critical-thinking instruction indicate that instructional design and learning activities require attention alongside the chosen tools (Lu & Xie, 2024; Shadiev& Wang, 2022). An Indonesian study of a mobile application offers a junior-secondary example of deliberately integrating reasoning activities, although its language, intervention, and outcome measures differ from this study (Agustina et al., 2022).

A digital resource would be relevant if it helped students revisit a text, compare evidence, or receive feedback on a justification. Digitizing a recall exercise would not change its cognitive demand. For MTsN 1 Pacitan, any proposed technology should be evaluated against access, teacher workload, and its contribution to the intended reasoning process. The central design decision remains the same for paper and digital materials: identify what learners must interpret or evaluate, provide sufficient language support, and collect a response that shows their reasoning.

Contribution and limits of the evidence

The study's main contribution is a situated account of connections among planning, teacher questioning, contextual tasks, learner participation, and assessment. It suggests that linguistic knowledge can become a resource for reasoning when students are asked to explain and evaluate meaning. However, triangulation strengthens the interpretation of a classroom process; it does not transform a qualitative case study into a causal-effect study. Methodological guidance

on case research and qualitative assessment supports maintaining a clear relationship between the research question, evidence, and scope of conclusions (Busetto et al., 2020; Rashid et al., 2019).

Several reporting limitations restrict interpretation. This report does not specify participant counts, observation duration, the precise fieldwork period, or a documented stopping criterion for data collection. This limits the transparency of the sampling and data-collection account. Research on saturation shows that adequacy depends on the research purpose and characteristics of the data, so a statement that "sufficient information" was obtained is not by itself a transparent justification (Hennink& Kaiser, 2022). Transparent reporting also requires an account of ethics and consent procedures, especially because the study involves school students.

The study also provides no baseline-to-follow-up measure, comparison group, independent rating of student work, or evidence of long-term transfer. These omissions do not invalidate its qualitative contribution, but they prevent claims about the size, persistence, or exclusive cause of improvement. A critical review of causal evidence in language education is relevant because it distinguishes promising instructional practice from demonstrated intervention effectiveness (El Soufi & See, 2019). Future work could combine detailed classroom analysis with clearly defined performance tasks and repeated assessment, while investigating how much reasoning students can demonstrate independently.

Finally, the comparative literature is predominantly drawn from English-language education. Its value lies in clarifying possible pedagogical processes, not in replacing evidence about Arabic or the Indonesian madrasah context. Future studies should examine whether the same task structures remain effective under Arabic-specific linguistic demands and different curricular conditions. The most defensible conclusion from this case is that HOTS-oriented practices create documented opportunities for interpretation and justification, while their consistency, inclusiveness, and measurable effects remain questions for further investigation.

Conclusions

The reported implementation of a HOTS-oriented Arabic curriculum at MTsN 1 Pacitan connects curriculum planning with contextual tasks, teacher questioning, discussion, and opportunities to explain answers. Its clearest contribution is the creation of occasions for students to interpret texts, compare alternatives, and justify a response. These practices suggest that language learning can provide a setting for critical reasoning when objectives, instructional support, and assessment attend to the quality of learners' explanations.

The case also shows why implementation must account for differences in Arabic proficiency and confidence. Linguistic support should help learners express their reasoning while retaining the intellectual demand of a task. Assessment should distinguish evidence of reasoning from linguistic accuracy and should not classify every comprehension or application question as HOTS. Professional collaboration can support teachers in reviewing these distinctions and improving the consistency of curriculum enactment.

The qualitative findings do not establish the magnitude, persistence, or causal source of improvement in critical thinking. A more complete account of participants, fieldwork, and analytic decisions is required, together with further evidence if claims about effectiveness are to be made. Future research should investigate Arabic-specific task demands, independent student performance, and transfer across lessons. The present study offers a contextual basis for that work without claiming effects beyond its reported evidence.

REFERENCES

Agustina, N., Mayuni, I., Iskandar, I., &Ratminingsih, N. M. (2022). Mobile learning application: Infusing critical thinking in the EFL classroom. *Studies in English Language and Education, 9*(2), 724–743. https://doi.org/10.24815/siele.v9i2.23476

Busetto, L., Wick, W., & Gumbinger, C. (2020). How to use and assess qualitative research methods. *Neurological Research and Practice, 2*(1), 14. https://doi.org/10.1186/s42466-020-00059-z

Cáceres, M., Nussbaum, M., & Ortiz, J. (2020). Integrating critical thinking into the classroom: A teacher's perspective. *Thinking Skills and Creativity, 37*, 100674. https://doi.org/10.1016/j.tsc.2020.100674

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

Double, K. S., McGrane, J. A., &Hopfenbeck, T. N. (2020). The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies. *Educational Psychology Review, 32*(2), 481–509. https://doi.org/10.1007/s10648-019-09510-3

El Soufi, N., & See, B. H. (2019). Does explicit teaching of critical thinking improve critical thinking skills of English language learners in higher education? A critical review of causal evidence. *Studies in Educational Evaluation, 60*, 140–162. https://doi.org/10.1016/j.stueduc.2018.12.006

Etemadfar, P., Soozandehfar, S. M. A., &Namaziandost, E. (2020). An account of EFL learners' listening comprehension and critical thinking in the flipped classroom model. *Cogent Education, 7*(1), 1835150. https://doi.org/10.1080/2331186x.2020.1835150

Gozali, I., Lie, A., Tamah, S. M., &Jemadi, F. (2021). HOTS questioning ability and HOTS perception of language teachers in Indonesia. *Indonesian Journal of Applied Linguistics, 11*(1), 60–71. https://doi.org/10.17509/ijal.v11i1.34583

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine, 292*, 114523. https://doi.org/10.1016/j.socscimed.2021.114523

Li, X., & Liu, J. (2021). Mapping the taxonomy of critical thinking ability in EFL. *Thinking Skills and Creativity, 41*, 100880. https://doi.org/10.1016/j.tsc.2021.100880

Li, L. (2023). Critical thinking from the ground up: teachers' conceptions and practice in EFL classrooms. *Teachers and Teaching, 29*(6), 571–593. https://doi.org/10.1080/13540602.2023.2191182

Liang, W., & Fung, D. (2021). Fostering critical thinking in English-as-a-second-language classrooms: Challenges and opportunities. *Thinking Skills and Creativity, 39*, 100769. https://doi.org/10.1016/j.tsc.2020.100769

Lintangsari, A. P., Emaliana, I., &Kusumawardani, I. N. (2022). Improving learners' critical thinking and learning engagement through Socratic questioning in Nominal Group Technique. *Studies in English Language and Education, 9*(2), 705–723. https://doi.org/10.24815/siele.v9i2.22352

Lu, D., & Xie, Y. N. (2024). Critical thinking cultivation in TESOL with ICT tools: a systematic review. *Computer Assisted Language Learning, 37*(1-2), 222–242. https://doi.org/10.1080/09588221.2022.2033788

Min, G. S., Mohd Ariff Albakri, I. S., Ismail, N., Mokhtar, M. M., Zulkepli, N., Tahir, M. H. M., & Khalid, P. Z. M. (2023). Fostering critical thinking using Graphic Organizers in English language reading class. *Studies in English Language and Education, 10*(3), 1309–1325. https://doi.org/10.24815/siele.v10i3.29973

Murtadho, F. (2021). Metacognitive and critical thinking practices in developing EFL students' argumentative writing skills. *Indonesian Journal of Applied Linguistics, 10*(3). https://doi.org/10.17509/ijal.v10i3.31752

Pelenkahu, N., Ali, M. I., Tatipang, D. P., Wuntu, C. N., &Rorintulus, O. A. (2024). Metacognitive strategies and critical thinking in elevating EFL argumentative writing proficiency: Practical insights. *Studies in English Language and Education, 11*(2), 873–892. https://doi.org/10.24815/siele.v11i2.35832

Purnamaningwulan, R. A. (2022). Listening journals to promote students' critical thinking skills in an integrated Listening-Speaking Course. *Studies in English Language and Education, 9*(2), 744–760. https://doi.org/10.24815/siele.v9i2.23387

Rashid, Y., Rashid, A., Warraich, M. A., Sabir, S. S., & Waseem, A. (2019). Case Study Method: A Step-by-Step Guide for Business Researchers. *International Journal of Qualitative Methods, 18*, 1609406919862424. https://doi.org/10.1177/1609406919862424

Sari, D. M. M., &Prasetyo, Y. (2021). Project-based-learning on critical reading course to enhance critical thinking skills. *Studies in English Language and Education, 8*(2), 442–456. https://doi.org/10.24815/siele.v8i2.18407

Shadiev, R., & Wang, X. (2022). A Review of Research on Technology-Supported Language Learning and 21st Century Skills. *Frontiers in Psychology, 13*, 897689. https://doi.org/10.3389/fpsyg.2022.897689

Soe, T. (2024). Investigating Critical Thinking in ELT Textbooks: A Systematic Literature Review of Textbook Evaluation Studies. *Theory and Practice of Second Language Acquisition, 10*(1), 1–29. https://doi.org/10.31261/tapsla.13882

Suharno, &Setyarini, S. (2021). Promoting EFL junior secondary students' critical thinking skills through analogical reasoning in narrative text. *Indonesian Journal of Applied Linguistics, 11*(1), 211–220. https://doi.org/10.17509/ijal.v11i1.34660

Susilo, A., Mufanti, R., & Fitriani, A. (2021). Promoting EFL students' critical thinking and self-voicing through CIRC technique in Academic Writing courses. *Studies in English Language and Education, 8*(3), 917–934. https://doi.org/10.24815/siele.v8i3.21149

Warsah, I., Morganna, R., Uyun, M., Hamengkubuwono, & Afandi, M. (2021). The Impact of Collaborative Learning on Learners' Critical Thinking Skills. *International Journal of Instruction, 14*(2), 443–460. https://doi.org/10.29333/iji.2021.14225a

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Frontiers in Psychology, 10*, 3087. https://doi.org/10.3389/fpsyg.2019.03087

Yan, Z. (2021). English as a Foreign Language Teachers' Critical Thinking Ability and L2 Students' Classroom Engagement. *Frontiers in Psychology, 12*, 773138. https://doi.org/10.3389/fpsyg.2021.773138

Yin, X., Saad, M. R. B. M., & Halim, H. B. A. (2023). A systematic review of critical thinking instructional pedagogies in EFL writing: What do we know from a decade of research. *Thinking Skills and Creativity, 49*, 101363. https://doi.org/10.1016/j.tsc.2023.101363

Yuan, R., & Stapleton, P. (2020). Student teachers' perceptions of critical thinking and its teaching. *ELT Journal, 74*(1), 40–48. https://doi.org/10.1093/elt/ccz044

Yuan, R., Liao, W., Wang, Z., Kong, J., & Zhang, Y. (2022). How do English-as-a-foreign-language (EFL) teachers perceive and engage with critical thinking: A systematic review

from 2010 to 2020. *Thinking Skills and Creativity, 43*, 101002. https://doi.org/10.1016/j.tsc.2022.101002

Yulian, R. (2021). The flipped classroom: Improving critical thinking for critical reading of EFL learners in higher education. *Studies in English Language and Education, 8*(2), 508–522. https://doi.org/10.24815/siele.v8i2.18366

Zhang, H., & He, Y. (2023). Facilitating language teachers' learning to teach critical thinking through lesson study. *Teachers and Teaching, 29*(4), 384–408. https://doi.org/10.1080/13540602.2023.2191179