From *Turāth* to Future Learning: Developing HOTS-Based Arabic Reading Materials from Sheikh Yusuf al-Makassari's Intellectual Legacy

Siti Maisaroh[1], Nurdiana Arifah[2], Alwiyah[3] and Sofiyatul Amaliyah[4]

[1]Arabic Language Education Program, Universitas Islam Negeri Madura, Pamekasan, Indonesia
sitimaisaroh@uinmadura.ac.id

Abstract

Arabic reading courses frequently devote sustained attention to vocabulary and grammar while offering students fewer opportunities to evaluate arguments, weigh evidence, and adapt inherited ideas to contemporary problems. Addressing this gap, the present educational design research developed and formatively evaluated a Higher-Order Thinking Skills (HOTS)–oriented *qirā'ah* (reading) unit grounded in the intellectual legacy of Sheikh Yusuf al-Makassari. The field-based development cycle encompassed three main phases: analysis and exploration, design and construction, and evaluation and reflection. Data collection involved twenty Arabic Language Education students who completed a needs analysis and a limited trial, semi-structured interviews with two *qirā'ah* lecturers, and expert reviews of the prototype. The needs analysis revealed the highest demand for culturally grounded content ($M = 4.62$), linguistic and cognitive scaffolding ($M = 4.55$), and future-oriented engagement ($M = 4.50$). These empirical insights informed the design of a six-stage pedagogical sequence (NUSANTARA–HOTS): Trace, Interpret, Connect, Evaluate, Transform, and Reflect. Expert validation yielded an overall score of 85.0%, student perception reached 83.9%, and rubric-based student performance across analysis, evaluation, and creation averaged 75.0%, with evaluation emerging as the most cognitively demanding task (73.8%). Lecturer testimony and qualitative student responses indicated that evaluating claims required learners to simultaneously coordinate Arabic text comprehension, textual evidence, evaluative criteria, and logical reasoning. Overall, the findings suggest that intellectual heritage becomes pedagogically productive when cultural relevance is aligned with source-status transparency, selective linguistic mediation, differentiated learning pathways, and evidence-based recontextualization. These results demonstrate the unit's formative feasibility rather than its definitive causal effectiveness.

Keywords: *Turāth*, Arabic reading, higher-order thinking skills, Sheikh Yusuf al-Makassari, educational design research.

© The Author(s) 2026
U. Faruq et al. (eds.), *Proceedings of the 10th International Conference of Islamic Studies (ICONIS 2026) & 400th Anniversary of Shekh Yusuf Al-Makassari*, Advances in Social Science, Education and Humanities Research 1043,
https://doi.org/10.2991/978-94-6239-845-0_13

1 Introduction

1.1 Background and Research Gap

Education in the face of uncertain futures requires learners to develop abilities that go beyond simply retrieving information. They are expected to engage critically with complex texts, examine and evaluate competing arguments, and apply existing knowledge responsibly when confronted with situations that cannot always be anticipated by their teachers. UNESCO describes this challenge as involving a renewed relationship between knowledge, technology, culture, and human agency [1]. The importance of this issue is also reflected in international reading assessments. According to the OECD, average reading performance experienced an unprecedented decline from 2018 to 2022 [2]. Indonesia encounters an even more challenging situation, as only around one in four participating students achieved Level 2 or above in reading in PISA 2022 [3]. Although these results do not directly represent the Arabic reading proficiency of university students, they nevertheless highlight a broader pedagogical concern: in what ways can reading instruction develop learners' capacity for critical judgment instead of merely enabling them to obtain information?

This issue is particularly relevant to Arabic Language Education, in which *qirā'ah* instruction has traditionally relied on activities such as explaining vocabulary, translating sentences, and analyzing grammatical structures. Although these approaches continue to play an important role, especially in helping students understand classical Arabic prose, they do not necessarily enable learners to identify the author's line of reasoning, examine contrasting ideas, or formulate judgments supported by evidence from the text. From a metacognitive perspective, effective reading is likewise understood as an intentional process involving planning, monitoring, and evaluating comprehension, rather than merely making a series of lexical decisions [4]. Recent developments in Arabic learning materials based on Higher-Order Thinking Skills (HOTS) have sought to move *qirā'ah* instruction beyond simple recall [5]. Related studies have also explored approaches to reading classical Islamic texts [6] as well as the development of interactive Arabic worksheets [7]. Nevertheless, many of these innovations continue to employ contemporary or broadly applicable content. As a result, the intellectual works produced by scholars of the Nusantara have rarely been positioned as a primary epistemic resource within Arabic reading instruction.

Recent findings provide further evidence of the significance of this issue. A systematic review published in 2024 on Arabic reading strategies among non-native learners shows that cognitive, metacognitive, affective, and social strategies function as complementary forms of support rather than as mutually interchangeable approaches [8]. Within the Indonesian context of Arabic education, the application of HOTS has also developed across several areas, including the production of instructional textbooks [5] the integration of technology and pedagogy [9] and the psychometric construction of C4–C6 assessment items [10]. Furthermore, a 2026 Sinta-2 study involving 21

undergraduate students found substantial pre–post improvements in critical Arabic reading, with an effect size of d = 0.99, as well as in reading self-efficacy, with an effect size of d = 1.64, following problem-based and value-integrated instruction [11]. Collectively, these studies demonstrate the importance of reading strategies, assessment practices, and cultural meaning in Arabic learning. However, they have not yet adequately addressed how the intellectual heritage of Arabic scholarship in the Nusantara can be transformed into a differentiated reading architecture in which the sources of knowledge remain transparent and traceable.

Sheikh Yusuf al-Makassari (1626–1699) represents a particularly valuable figure for this purpose. His intellectual journey extended across the Indonesian archipelago, Arabia, Sri Lanka, and South Africa, and his writings brought together intellectual knowledge, spiritual formation, ethical commitment, and courage in responding to social realities [12,13]. The enduring significance of his Arabic intellectual heritage is also reflected in the philological edition of *Zubdat al-Asrār* produced by Nabilah Lubis [14]. Nevertheless, existing scholarship has predominantly examined Sheikh Yusuf from historical, biographical, doctrinal, or political perspectives. Comparatively little attention has been given to the pedagogical potential of his intellectual legacy, particularly to the ways in which his Arabic writings could be adapted into meaningful and teachable reading experiences for Arabic language learners.

Taken together, these issues reveal three interconnected research gaps. The first concerns content, as Arabic reading materials used in Indonesian higher education seldom position Arabic works produced within the Nusantara as central sources of knowledge. The second gap is pedagogical in nature. Although classical Arabic texts are commonly taught through translation and explanatory activities, they are less frequently transformed into learning tasks that require learners to analyze evidence, evaluate arguments, and generate new interpretations. The third gap relates to development, since research on Sheikh Yusuf's intellectual heritage and research on HOTS-based *qirā'ah* have generally developed independently rather than being integrated into a unified instructional framework. Merely pairing a historical Arabic passage with a set of challenging questions would therefore be insufficient to address these three gaps. A more systematic and traceable design is required—one that maintains clear epistemic boundaries, provides appropriate mediation for linguistic complexity, and progressively organizes the level of cognitive demands placed on learners.

1.2 Research Questions and Contribution

The present study responds to these needs by developing and conducting a formative evaluation of the proposed design. It addresses three main questions: (1) what kinds of learning needs do students and lecturers identify for HOTS-oriented Arabic reading materials based on the intellectual heritage of Sheikh Yusuf al-Makassari; (2) in what ways can this intellectual heritage be systematically adapted into a *qirā'ah* prototype that engages learners in analysing, evaluating, and creating; and (3) how valid and pedagogically feasible is the resulting prototype based on expert assessment and a limited trial involving students? The principal contribution of the study is the

NUSANTARA–HOTS architecture, which brings together *turāth* as an epistemic foundation, *qirā'ah* as a means of linguistic engagement, HOTS as a framework for cognitive activity, and future-oriented learning as a process of responsible recontextualization. The investigation is limited to a single formative development cycle and, therefore, does not seek to establish causal claims regarding the effectiveness of the prototype.

1.3 Conceptual Framework

The framework is built upon four interrelated propositions. First, reading is understood as a regulated process of meaning construction in which learners deliberately plan their reading, identify and monitor difficulties in comprehension, choose suitable strategies, and assess the interpretations they formulate [15]. Recent research on scaffolding indicates that the use of structured visual and metacognitive support can help reduce unnecessary cognitive load while encouraging learners to move beyond surface-level recall toward deeper inferential processing [16,17]. Second, effective scaffolding needs to be contingent on learners' needs and responses. Such support is most productive when it makes expert thinking processes explicit, provides cues that help diagnose difficulties, and is progressively reduced as learners become more capable, rather than continuously simplifying the text itself [17–19].

Third, HOTS in this framework is understood as performance that is grounded in evidence rather than being determined merely by the labels assigned to questions. At the analysis level, learners examine relationships among ideas while identifying the textual evidence that supports those relationships. Evaluation requires learners to bring together a claim, relevant evidence, appropriate criteria, a logical warrant, and an alternative or counter-position. At the creation level, learners are expected to transform a defensible interpretation into a product within clearly defined boundaries [20]. Consistent with this perspective, recent developments in Arabic HOTS assessment have operationalized C4–C6 by considering evidence related to content, construct, language, level of difficulty, and item discrimination [21]. Fourth, culturally grounded pedagogy demands epistemic accountability. Cultural representation becomes educationally meaningful when learners can distinguish which ideas originate from the source, which portions have been paraphrased, and which elements represent interpretations developed in the present context [22].

These propositions establish the study's design-level causal logic, without implying a claim about causal effects. The process begins with identified learner needs, which guide decisions concerning the selection of sources and the mediation of linguistic complexity. This mediation then supports the implementation of six reading operations, while evidence obtained from learning tasks and assessment rubrics is used to inform formative revisions. The resulting revised design is subsequently used to derive principles that may be transferable to similar instructional contexts. This sequence of reasoning is presented explicitly in Figure 1.

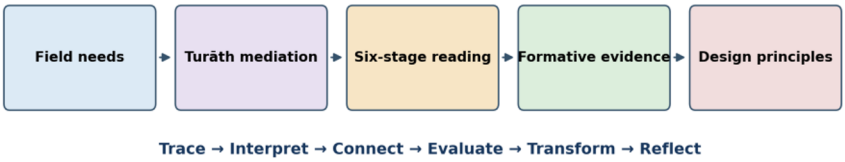


Fig. 1. The Evidence-to-design logic of the NUSANTARA–HOTS prototype.

2 Method

2.1 Design and Setting

The study used educational design research (EDR), which combines systematic inquiry with the iterative construction of educational solutions in authentic settings [23,24]. It was conducted in the Arabic Language Education Program at Universitas Islam Negeri Madura, Indonesia. One formative cycle followed three connected phases: analysis and exploration, design and construction, and evaluation and reflection [25]. This scope was appropriate because the aim was to produce an initial pedagogical prototype and generate design knowledge, rather than to estimate an intervention effect.

Participants were selected purposively because of their direct relation to *qirā'ah* instruction. Twenty undergraduate students participated in both the needs analysis and the limited trial; thus, the study involved 24 individuals rather than two separate student cohorts. Ten students were in the fourth semester and ten in the sixth. Existing course records and lecturer judgment placed five students in the high, eleven in the moderate, and four in the low Arabic-proficiency band. These bands were used only to plan differentiated support and describe responses; no inferential subgroup comparison was attempted. Two lecturers had experience teaching Arabic reading, while the two validators represented Arabic reading/materials expertise and HOTS-oriented instructional design.

2.2 Procedure and Prototype Construction

The fieldwork took place from 5 to 7 August 2026. At the analysis and exploration stage, students responded to an eleven-item needs-assessment questionnaire using a five-point scale. The questionnaire focused on their existing reading practices, previous experience with HOTS, the linguistic assistance they required, cultural identity, and engagement with future-oriented learning. Interviews with lecturers were conducted to investigate the strengths and limitations of the materials currently in use, the practicality of incorporating classical Arabic texts, and the conditions needed to support students in progressing from textual interpretation toward reasoned judgment. The patterns

identified from the questionnaire responses and the codes generated from the lecturer interviews were then translated into specific requirements for the instructional design.

During the design stage, an 8–12-page Arabic reading unit was developed under the title العِلْمُ وَالأَخْلَاقُ وَمَسْؤُولِيَّةُ الإِنْسَانِ فِي بِنَاءِ الْحَضَارَةِ (*Knowledge, Ethics, and Human Responsibility in Building Civilization*). The unit presented an adapted pedagogical Arabic text constructed around themes drawn from the intellectual legacy associated with Sheikh Yusuf, including knowledge, moral discipline, self-cultivation, social responsibility, intellectual mobility, and resistance to injustice. The material was intentionally not framed as a verbatim transcription of a manuscript or as a critical edition of an original text. Instead, the design maintained a clear distinction between claims derived from primary sources, scholarly paraphrases, pedagogical adaptations, and interpretations situated within contemporary contexts.

The learning activities were designed through constructive alignment with the intended reading outcomes and the accompanying analytic rubric. The instructional sequence—Trace, Interpret, Connect, Evaluate, Transform, and Reflect—progresses from establishing an understanding of the source and providing access to its linguistic content, toward examining relationships among ideas, developing evidence-based judgments, producing a contextually bounded product, and engaging in metacognitive reflection. To accommodate differences in learners' proficiency levels, the design also incorporates Foundation and Challenge routes, allowing students to receive different levels of support while maintaining the same overarching intellectual learning goal.

2.3 Instruments, Analysis, and Quality Safeguards

Four instruments were used: an eleven-item student needs questionnaire, a six-question semi-structured lecturer interview guide, a 12-indicator expert-validation sheet, and a ten-item student-response questionnaire accompanied by an analytic performance rubric. Closed items used five-point scales. Student work was scored on five four-point dimensions: analysis, evaluation, creation, Arabic language, and reflection. For each scale, the observed score was divided by the maximum possible score and multiplied by 100. Percentages of 81–100 were interpreted as very feasible, 61–80 as feasible, 41–60 as moderately feasible, 21–40 as less feasible, and 0–20 as not feasible. Given the small purposive group, all analyses remained descriptive.

Descriptive analysis included item and dimension means, standard deviations, medians, observed ranges, 95% confidence intervals for sample means, and the proportion selecting 4 or 5. Internal consistency was estimated only for scales intended to represent a common construct. The ten-item post-use response scale showed high consistency (Cronbach's $\alpha = .926$), as did the five-dimension performance rubric ($\alpha = .951$). The multidimensional needs instrument was not treated as a single reflective scale because its dimensions intentionally sampled distinct design requirements; therefore, a total-scale alpha was not used as evidence of validity. Proficiency-band results are descriptive because the groups ($n = 5$, 11, and 4) were small and formed from existing records rather than random assignment.

The researcher scored the performance tasks using the analytic rubric after students had completed the unit. Because a second independent rater was not available in this

cycle, no inter-rater coefficient is reported; this is treated as a limitation rather than concealed by the precision of the percentages. Lecturer interviews produced 12 meaning units. The researcher condensed each unit, assigned an initial code, compared codes across lecturers, and grouped them into six themes: literal-reading dominance, *turāth* as a living source, linguistic scaffolding, evidence-based argumentation, humane future learning, and proficiency differentiation. Negative cases were retained when a lecturer challenged a uniform progression or excessive glossing. Open student responses and validator notes were analysed using the same code–category–theme sequence.

Triangulation examined convergence among survey patterns, lecturer accounts, expert judgment, student response, and observed task performance. An audit trail linked each major design feature to an identified need, theoretical rationale, validator request, and revision. All 20 students recorded consent before the limited trial. Participants were informed of the study purpose, voluntary participation, anonymity, and intended scholarly use of the data. Student and lecturer identities are reported only through codes. The two-validator score is treated as formative design evidence, not as a psychometric claim.

The prototype is a thematic pedagogical adaptation rather than a critical edition. It distinguishes source-based ideas, scholarly paraphrase, pedagogical reformulation, and present-day interpretation. No sentence in the adapted reading is claimed as a verbatim manuscript quotation without page- or folio-level verification.

3 Results and Discussion

3.1 Needs: Critical Reading Requires More Than Difficult Questions

The needs assessment indicated that students were not simply asking for more difficult exercises. The dimension related to context and intellectual identity recorded the highest mean score (M = 4.62), followed by scaffolding (M = 4.55), future learning and engagement (M = 4.50), and HOTS (M = 4.02). At the individual-item level, the perceived importance of Nusantara scholars in shaping the intellectual identity of PBA students obtained a mean of M = 4.75. Both contextual connection and the provision of a contextual glossary received high mean scores of M = 4.65. By comparison, the reverse-scored item concerning students' ability to independently analyse relationships among ideas produced a lower mean of M = 3.35, with eleven of the twenty respondents assigning scores of four or five.

Across the twelve scored needs indicators, the average score at the participant level was 4.41 (SD = 0.16, median = 4.42), with observed scores ranging from 4.17 to 4.67 and a 95% confidence interval of [4.34, 4.48]. Eleven out of the twenty students (55%) rated their ability to independently analyse relationships among ideas at either 4 or 5. In comparison, the other major forms of support received endorsement from 18 to 20 students (90–100%). This pattern suggests that the findings do not reflect a broad problem of low motivation. Rather, they point specifically to independent analytical

control as the principal area of weakness amid an otherwise strong demand for the identified learning supports.

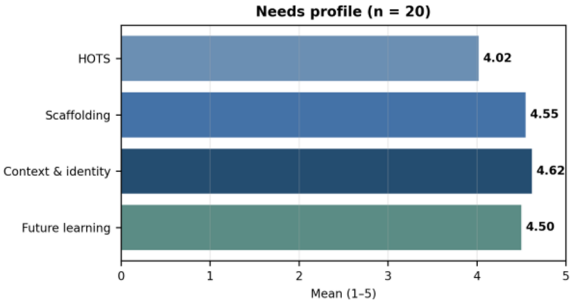


Fig. 2. The Mean student needs by design dimension (scale 1–5).

Table 1. The student needs by design dimension (n = 20)

Dimension	No.	M	Design implication
HOTS	3	4.02	High need; independent analysis remains uneven
Scaffolding	3	4.55	Glossary, selective vowel marking, and guided evidence use
Context and identity	3	4.62	Nusantara scholarship as meaningful content
Future learning and engagement	2	4.50	Contemporary connection with human agency

The open responses clarified the pattern. Students distinguished being able to translate from being able to explain why an author advanced a claim. One moderate-proficiency student wrote, “Give us the option to answer the evaluation in Indonesian first and then summarise it in Arabic” (S02). A low-proficiency student identified “the long sentence in the second paragraph” as the main difficulty and requested a sentence map (S04). Stronger readers, by contrast, wanted less glossing and more argumentative challenge. Willingness to engage with HOTS was therefore stronger than confidence in performing it without mediation.

The lecturer interviews converged with this diagnosis. “When students are asked to explain relationships among ideas, they often return to translating the sentence,” observed one lecturer (L1). The second added that many tasks still stop at “who, what, and translate,” leaving students little practice in defending an interpretation (L2). Both lecturers resisted removing vocabulary and grammar; their concern was the assumption that linguistic difficulty itself constituted higher-order thinking. L1 argued that evaluation was realistic only when students were shown how to use textual evidence, while L2 recommended guided questions and a bounded product rather than a long Arabic essay. Two negative cases were consequential: moving every student immediately to creation would overload weaker readers, while excessive glossing would flatten the experience of stronger readers.

Table 2. The Joint display of field evidence and design response

Theme	Concrete evidence	Design consequence
Literal dominance	11/20 endorsed independent analysis; both lecturers described return to translation	Worked C4 map; evidence prompts
Linguistic load	Long sentences, abstract Sufi terms, iḍāfah, and reasons in Arabic recurred	Selective ḥarakāt, glossary, sentence map
Evidence use	"Choosing the strongest evidence" was named as hardest	Claim–evidence–reasoning example and rubric
Differentiation	5 high, 11 moderate, 4 low; lecturers rejected one-size progression	Foundation and challenge routes
Source status	Students and validators requested source transparency	Labels for source, paraphrase, adaptation, interpretation

This finding complicates the familiar solution of relabeling questions as C4, C5, or C6. Teachers may endorse higher-order thinking while still facing constraints of language readiness, task design, and classroom time [26]. In the present study, students did not lack interest; they lacked a visible method for thinking with a text. The design problem was therefore ecological: context, linguistic access, evidence models, criteria for judgment, and opportunities for production had to work together.

3.2 The NUSANTARA–HOTS Design

NUSANTARA is used in this study as a conceptual designation rather than as an acronym derived from the English names of the instructional stages. The term signifies the grounding of the proposed architecture in the Arabic intellectual heritage of scholars from the Nusantara [27]. Within this study, future learning is operationally understood as the learner's ability to engage with inherited knowledge, assess its relevance by drawing on evidence from the text, and formulate an ethically responsible response when confronted with an unfamiliar contemporary issue.

The six-stage sequence operationalizes these requirements into concrete and observable reading activities. The Trace stage builds learners' source literacy and provides historical orientation, while interpret integrates linguistic scaffolding with inferential comprehension. Connect then develops relational reasoning and encourages learners to transfer their understanding to relevant contexts. In the Evaluate stage, learners construct judgments through a claim–evidence–reasoning framework and are required to consider a counterpoint. Transform enables them to reconstruct and apply knowledge through the production of a new, clearly bounded product. Finally, Reflect strengthens metacognitive monitoring and the development of ethical agency [28]. Although the stages are deliberately sequenced, the process is not strictly linear. Learners may move back from evaluation to interpretation, particularly when they discover that the evidence or its meaning has been misunderstood.

Table 3. The NUSANTARA–HOTS instructional sequence

Stage	Purpose	Focus	Learning evidence
Trace	Locate source, context, and concepts	Understand	Context hypothesis and source identification
Interpret	Construct linguistic and contextual meaning	Apply–analyse	Paraphrase and inference
Connect	Relate ideas and a present issue	Analyse	Relationship map
Evaluate	Judge strength, relevance, and limits	Evaluate	Claim, evidence, reason, and counterpoint
Transform	Reconstruct an idea for a new context	Create	Short recommendation or infographic
Reflect	Examine ethics and one's thinking	Reflect	Metacognitive response

Three task types made the cognitive progression visible. At C4, students mapped how knowledge, ethics, and social responsibility were connected and cited the sentence supporting each link. At C5, they judged whether a contemporary educational practice was consistent with the adapted text, using a claim, two pieces of evidence, a warrant, and one counterargument. At C6, they produced a 150-word recommendation or infographic that transferred the text's ethical reasoning to a new learning problem while identifying which elements were textual evidence and which were their own interpretation.

The sequence rests on a distinction often blurred in Arabic reading: linguistic difficulty and cognitive complexity are related but not equivalent. Translating an obscure sentence may be linguistically demanding while remaining a retrieval or application task. Conversely, evaluating two accessible claims and defending a judgment may require substantial higher-order reasoning [29]. NUSANTARA–HOTS therefore uses selective linguistic mediation to free cognitive capacity for the intended intellectual work, not to simplify away the text's conceptual character.

Its cultural grounding also requires epistemic discipline. Sheikh Yusuf's name cannot serve as an ornamental label for contemporary claims. The prototype treats present-day issues including artificial intelligence as researcher-led recontextualizations, not ideas attributed anachronistically to a seventeenth-century author. This responds to concerns that culturally responsive pedagogy can become symbolic when representation is detached from serious engagement with knowledge and context [26]. In the current cycle, the absence of a completed page- or folio-level audit limits the strongest possible claim: the unit is a thematic pedagogical adaptation of Sheikh Yusuf's legacy, not a critical edition or direct reconstruction of his manuscript language.

3.3 Expert Review: Feasibility with Consequential Revisions

Across 24 ratings from two experts, the prototype obtained an overall mean of 4.25/5, Based on 24 ratings provided by two experts, the prototype achieved an overall mean score of 4.25 out of 5, equivalent to 85.0%. The highest evaluations were given to the aspects of scaffolding, constructive alignment, recontextualization, future orientation, and the naturalness of the Arabic language used. Meanwhile, source fidelity, readability, and usability each received a score of 80.0%. These results indicate that, despite the prototype attaining a strong overall evaluation, several aspects still required further refinement.

Table 4. The Expert validation and design decisions (n = 2)

Aspect	M	%	Revision decision
Source fidelity	4.00	80.0	Clarify quotation, paraphrase, adaptation, and source status
Arabic language/readability	4.25	85.0	Correct iḍāfah; add selective ḥarakāt and segmentation
Scaffolding/alignment	4.50	90.0	Retain sequence; add worked example
HOTS alignment	4.17	83.3	Differentiate support and bound creation task
Recontextualization /future learning	4.50	90.0	Retain with explicit anti-anachronism labels
Usability	4.00	80.0	Simplify instructions and extend reading time
Overall	4.25	85.0	Very feasible after targeted revision

The written feedback produced concrete changes. Epistemic labels were added; two iḍāfah constructions were corrected; selective vowel marking and sentence mapping were introduced; a worked example modelled claim–evidence–reasoning; foundation and challenge routes were made explicit; and the creation task was limited to an infographic or a 150-word recommendation. In EDR, this trace between evidence and revision matters more than the percentage alone because it shows how the solution changed in response to use and judgment[23,30].

3.4 Limited Trial: Acceptance Is Not Mastery

The limited student trial demonstrated a generally favorable response, with an overall mean score of 4.20 out of 5 (83.9%). Recontextualization emerged as the most highly rated aspect, reaching 94.0%, followed by students' interest in Nusantara scholars at 92.0% and scaffolding at 90.0%. In contrast, reasoned evaluation obtained the lowest score at 79.0%. This distribution indicates that the prototype's cultural and intellectual relevance, together with the accessibility provided through its supports, constituted immediate strengths, whereas developing the capacity for disciplined and well-reasoned judgment continued to present a greater challenge for students.

Table 5. The Selected student-response and performance results (n = 20)

Evidence	Component	M	%
Response	Scaffolding	4.50	90.0
Response	Reasoned evaluation	3.95	79.0
Response	Recontextualization	4.70	94.0
Response	Overall	4.20	83.9
Rubric	Analysis	3.05	76.3
Rubric	Evaluation	2.95	73.8
Rubric	Creation	3.00	75.0
Rubric	Reflection	3.65	91.3
Rubric	Core HOTS	3.00	75.0

Rubric evidence tempers the response score. Analysis reached 76.3%, creation 75.0%, and evaluation 73.8%; the three core HOTS components averaged 75.0%. Reflection was much higher at 91.3%. That contrast should not be read as proof that reflection was the most advanced cognitive achievement. Personal reflection is close to lived experience and may not require the same evidential discipline as evaluation. Evaluation simultaneously demands comprehension of the author's claim, selection of criteria, identification of relevant evidence, and a justified decision—all while the student manages Arabic vocabulary and syntax.

The participant-level core HOTS score was M = 3.00/4 (SD = 0.61, median = 3.00, range = 2.00–4.00; 95% CI [2.71, 3.29]). Component dispersion was similarly visible: analysis M = 3.05 (SD = 0.69), evaluation M = 2.95 (SD = 0.61), creation M = 3.00 (SD = 0.65), Arabic language M = 3.05 (SD = 0.69), and reflection M = 3.65 (SD = 0.49). The response-scale mean was 4.20/5 (SD = 0.48, median = 4.30, range = 3.40–4.90; 95% CI [3.97, 4.42]). Thus, the response–performance gap was present not only in percentages but also in dispersion: students differed substantially in what they could perform after generally accepting the unit.

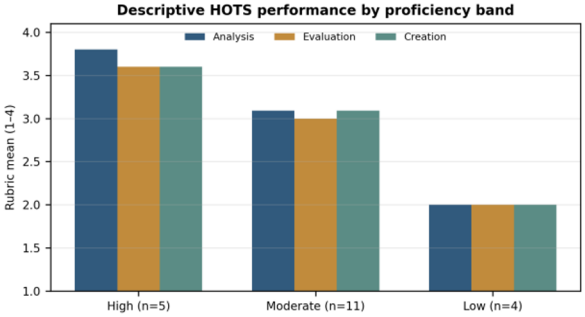


Fig. 3. The Descriptive HOTS performance by prior Arabic-proficiency band. No inferential comparison was conducted.

Table 6. The Descriptive results by prior Arabic proficiency

Band (n)	Response	Analysis	Evaluation	Creation	Core HOTS
High (5)	4.70	3.80	3.60	3.60	3.67

Moderate (11)	4.25	3.09	3.00	3.09	3.06
Low (4)	3.40	2.00	2.00	2.00	2.00

The band pattern was monotonic: core HOTS means were 3.67 for high-, 3.06 for moderate-, and 2.00 for low-proficiency students. Because the bands were small and predetermined, this is diagnostic evidence, not an inferential group effect. It nevertheless supports a concrete design decision: low-proficiency learners require an extended Interpret route and bilingual planning before evaluation, while high-proficiency learners require reduced glossing and stronger counterargument demands.

The 8.9-point gap between perceived feasibility (83.9%) and core HOTS performance (75.0%) is analytically useful. Students accepted the unit and recognized its value, but acceptance did not equal mastery. High-, moderate-, and low-proficiency students also showed descriptively different experiences, reinforcing the need for foundation and challenge routes. Research on personalized and metacognitive scaffolds likewise stresses contingency, diagnostic cues, and gradual transfer of responsibility [18,19,31]. Direct modelling and inquiry are therefore complementary rather than competing choices [32].

Open responses identified the contextual introduction, glossary, concept map, peer discussion, guided questions, and worked example as useful. One student named "choosing the strongest evidence" as the hardest step (S05), while another asked for "one example answer showing claim, evidence, and reason" (S01). Difficulties also clustered around abstract Sufi terminology, long sentences, separating source information from interpretation, and expressing reasons in Arabic. These data make the next design priority concrete: paired examples of strong and weak evaluation, a visible claim–evidence–reasoning matrix, gradual prompt reduction, and an analytic rubric separating comprehension, evidence relevance, reasoning quality, and depth of judgment.

All 20 students supplied three open responses, yielding 60 short field statements. Content coding produced four recurrent difficulty clusters: linguistic–conceptual load (long syntax, abstract vocabulary, metaphor, iḍāfah, and Arabic formulation), evidence and epistemic control (selecting evidence, remaining within the text, and separating fact from interpretation), relational reasoning (main–supporting ideas, cross-paragraph links, and causal explanation), and product regulation (summary, recommendation, critique, and time). The improvement requests were correspondingly concrete: one worked claim–evidence–reasoning answer; paired strong/weak evidence; marginal notes for Sufi terms; a sentence map; a fact-versus-interpretation column; a basic route; oral or Indonesian-first planning options; and approximately ten additional reading minutes.

3.5 From Heritage Preservation to Future Learning

Four forms of transformation illustrate how *turāth* was made pedagogically active within the prototype. Epistemic transformation clarifies the status and origin of the material presented in each passage. Linguistic transformation provides access to challenging content while preserving the essential Arabic concepts embedded within

the text. Cognitive transformation reorients reading activity from simple information retrieval toward evidence-based analysis, evaluation, creation, and reflection. Participatory transformation, meanwhile, provides learners with differentiated learning routes, opportunities for peer interaction, and clearly bounded spaces in which they can formulate and present their own responses [33]. Each transformation serves a distinct function, but none is adequate when implemented in isolation.

The proposed model builds upon earlier developments in HOTS-oriented *qirā'ah* [5] by demonstrating that higher-order questioning cannot function effectively without an underlying structure of evidence, scaffolding, and explicit evaluative criteria. Worked examples can initially expose learners to the processes involved in evidence-based reasoning before such support is gradually withdrawn through fading [34], Research on reading strategies likewise emphasizes the importance of integrating comprehension monitoring, learner motivation, and explicit strategy instruction [35,36]. In relation to scholarship on Sheikh Yusuf [8–10], the model moves beyond historical interpretation by translating his intellectual legacy into pedagogically operational forms while maintaining a clear boundary against anachronistic interpretation. Its context-sensitive and iterative approach is also consistent with perspectives that emphasize the importance of context and implementation, rather than treating educational evidence as independent of its setting [37]. Similarly, culturally sustaining literacy encourages learners to actively extend and develop their cultural resources rather than merely reproduce or demonstrate them [37]. Within this framework, future learning should not be reduced to the incorporation of digital technologies or generative AI. In accordance with human-centred educational guidance [1,38], it refers instead to learners' capacity to critically, ethically, and productively apply inherited knowledge when responding to new and unfamiliar problems.

The pattern observed in the fieldwork is also consistent with more recent comparative findings. Mekheimer's 2025 study on digital scaffolding reported that a structured before–during–after reading organizer generated a large adjusted effect on comprehension (partial $\eta^2 = .667$) and encouraged deeper inferential responses [30]. The present formative cycle does not provide sufficient evidence to claim an effect of comparable magnitude; instead, it points toward a similar underlying mechanism, namely that making the structure of the reading process explicit can reduce the coordination demands placed on learners. In a related vein, Eticha et al. [21] highlight the role of metacognitive scaffolding in contextualized problem solving, while Setiadi et al. [16] demonstrate that Arabic texts with meaningful cultural content can contribute to both critical reading and reading self-efficacy. The NUSANTARA–HOTS architecture extends this body of work by incorporating transparent identification of source status and an explicit anti-anachronism boundary. These features are particularly important when instructional materials are constructed from the intellectual legacy of a specific historical scholar.

The central design proposition can be expressed as follows: Arabic reading grounded in cultural contexts becomes oriented toward future learning when its intellectual themes undergo source verification, its linguistic complexity is appropriately mediated, its cognitive demands are systematically sequenced, and its activities are differentiated according to learners' levels of readiness. At the same time, learners should be given

opportunities to reconstruct and apply these ideas to contemporary contexts while maintaining accountability to evidence contained in the original texts. This proposition represents the principal transferable outcome generated through the formative cycle, whereas the prototype itself remains context-specific and formative in nature.

3.6 Boundaries and Next Cycle

The study was conducted within a single program and involved two lecturers, two expert validators, 20 students, one instructional unit, and one development cycle. The trial generated evidence concerning users' perceptions and students' task performance rather than evidence of pre–post changes. Since the study did not include a control group or comparison condition, and the small purposively selected proficiency groups were not suitable for inferential comparisons, the findings should be interpreted within these methodological limitations. Accordingly, the study supports claims of preliminary validity, feasibility, learner engagement, and the elicitation of HOTS-related performance. However, it does not provide sufficient evidence to conclude that the prototype itself produced measurable improvement, nor can the findings be generalized to Arabic Language Education programs in broader contexts.

The subsequent development cycle should undertake a more comprehensive audit of primary sources at the page and folio levels, increase the number of instructional units, and involve at least two trained raters to strengthen the assessment process. Future research should also investigate whether the learning outcomes can be transferred to unfamiliar Arabic texts. In addition, the two scaffolding routes should be systematically compared, supported by pre–post measurements involving a larger and more diverse sample of learners. If digital or AI-assisted annotation is incorporated into the design, its contribution should be assessed on the basis of accuracy, proper source attribution, learner agency, and academic integrity, rather than being valued simply because of its technological novelty.

4 Conclusion

This study began with a practical problem: students could often translate Arabic passages but had fewer structured opportunities to reason with them. The needs data showed that the strongest demand was not for cognitive difficulty in isolation. Students wanted culturally meaningful content, linguistic and cognitive scaffolding, and a route from inherited ideas to contemporary questions. Those requirements shaped a single NUSANTARA–HOTS unit organized through Trace, Interpret, Connect, Evaluate, Transform, and Reflect.

Expert validation (85.0%) and student response (83.9%) indicate strong formative feasibility after targeted revision. Core HOTS performance reached 75.0%, with evaluation emerging as the central bottleneck. The result is encouraging but precise: the prototype elicited analysis, evaluation, creation, and reflection in a limited setting; it did not establish causal effectiveness. Its most defensible contribution is a design

logic that joins source transparency, selective linguistic mediation, differentiated scaffolding, evidence-based judgment, and responsible recontextualization.

Moving from *turāth* to future learning is therefore not a departure from the past. It is a disciplined conversation with it. Sheikh Yusuf's legacy becomes educationally alive when students are helped to locate a source, understand its language, connect its ideas, test its relevance, reconstruct its implications, and reflect on their own intellectual responsibility. Completing the source audit and strengthening evaluative argumentation are the necessary next steps before the model can be tested at scale.

Acknowledgments. The authors would like to express their gratitude to the Arabic Language Education Program at Universitas Islam Negeri Madura for supporting the implementation of this study. The authors also extend their appreciation to the participating lecturers, expert validators, and students who contributed to the needs analysis, prototype evaluation, and limited trial. Their participation and constructive feedback were valuable in the formative development of the NUSANTARA–HOTS Arabic reading materials.

Disclosure of Interests. The authors declare that they have no competing interests related to this study or its publication.

References

1. Carney S. Reimagining our futures together: a new social contract for education: by UNESCO, Paris, UNESCO, 2021, 186 pages, ISBN 978-92-3-100478-0. *Comparative Education*. 2022;58(4):568-569. doi:10.1080/03050068.2022.2102326
2. OECD. *PISA 2022 Results (Volume I) The State of Learning and Equity in Education: The State of Learning and Equity in Education*. OECD Publishing; 2023. doi:10.1787/53f23881-en
3. OECD. *PISA 2022 Results: Country Note—Indonesia*. OECD Publishing; 2023. doi:https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/indonesia_c2e1ae0e-en.html
4. Razkane H, Sayeh AY, Diouny S, Yeou M. Eleventh-grade Students' Use of Metacognitive Reading Strategies in Arabic (L1) and English (L3). *INT J INSTRUCTION*. 2023;16(1):573-588. doi:10.29333/iji.2023.16132a
5. Nashoih AK. HOTS-Based Qiro'ah Textbook Development to Enhance Arabic Learners' Extensive Reading Skills. *Dinamika*. 2024;9(2):24-42. doi:10.32764/dinamika.v9i2.4751
6. Syafei I, Suleman E, Rohanda R. The Development of Student Reading Skills in Arabic for Reading Islamic Classical Books Using the Arabic Learning Model at Indonesian Islamic Boarding Schools. *tpls*. 2024;14(5):1381-1392. doi:10.17507/tpls.1405.10
7. Zizah R. Empowering Arabic Reading Skills through Interactive Digital Worksheets: A Development Study Using Liveworksheet. *Alibbaa': Jurnal Pendidikan Bahasa Arab*. 2025;6(2):309-329. doi:10.19105/ajpba.v6i2.20923

8. Nash BL, Dunham H, Murdter-Atkinson J, Mosley Wetzel M. A Culturally Sustaining Approach to Multimodal Literacies: Building from Students' Funds of Knowledge in Field-Based Teacher Education. *Literacy Research and Instruction*. 2024;63(2):151-175. doi:10.1080/19388071.2022.2153766
9. Saepudin S, Pabbajah MTH, Pabbajah M. Unleashing the Power of Reading: Effective Strategies for Non-Native Arabic Language Learners. *ALSINATUNA*. 2024;9(2):109-130. doi:10.28918/alsinatuna.v9i2.7826
10. Ahmadi Y, Triana HW, Ritonga M. TPACK approach to improving HOTS in Arabic language learning: Strategies and challenges. *Shaut al Arabiyyah*. 2024;12(2):360-370. doi:10.24252/saa.v12i2.44118
11. Masrifah K, Ahsanuddin M. Development of HOTS-Based Final Semester Assessment Questions in the Al-'Ashri Arabic Textbook for 10th-grade Students of Islamic Senior High School. *JBA*. 2025;9(2):755-769. doi:10.29240/jba.v9i2.13889
12. Arif S. Shaykh Yūsuf of Makassar (d. 1111 AH/1699 CE): A Bio-bibliographical and Doctrinal Survey. *Intellectual Discourse*. 2024;32(1). doi:10.31436/id.v32i1.2079
13. Umam ZK. The Scribe of Sufi-Philosophical Letters: Shaykh Yūsuf of Makassar's Formative Decades (1640s-1660s) in Arabia and Syria. *Stud Islam*. 2024;31(3):405-438. doi:10.36712/sdi.v31i3.42953
14. Lubis N. *Menyingkap Intisari Segala Rahasia: Syekh Yusuf al-Taj al-Makasari*. Fakultas Sastra, Universitas Indonesia; 1996. doi:https://books.google.co.id/books/about/Menyingkap_intisari_segala_rahasia.html?id=W9jjAAAAMAAJ&redir_esc=y
15. Guzmán Murillo HJ, Torres Ortega JM, González Sánchez DA. PlanSuperEva: Estrategias Metacognitivas para el Fortalecimiento de la Lectura Crítica en Educación Secundaria y Media. *Ciencia Latina*. 2024;8(5):13090-13104. doi:10.37811/cl_rcm.v8i5.14765
16. Setiadi S, Mahmudah I, Hanifah R, Mutmainah M, Nafiah K, Masyitoh S. The Effectiveness of Problem-Based Critical Arabic Reading Integrating Nabawi Peace Values on Students' Reading Achievement and Self-Efficacy. *Eralingua*. 2026;10(1). doi:10.26858/eralingua.v10i1.83066
17. Mekheimer M. Digital mind magnets as a scaffold for improving EFL online reading comprehension. *Discov Educ*. 2025;4(1):531. doi:10.1007/s44217-025-00988-4
18. Kärner T, Warwas J, Schumann S. A Learning Analytics Approach to Address Heterogeneity in the Classroom: The Teachers' Diagnostic Support System. *Tech Know Learn*. 2021;26(1):31-52. doi:10.1007/s10758-020-09448-4
19. Van Der Graaf J, Raković M, Fan Y, et al. How to design and evaluate personalized scaffolds for self-regulated learning. *Metacognition Learning*. 2023;18(3):783-810. doi:10.1007/s11409-023-09361-y
20. Abbas A, Hefnawy MT, Negida A. Meta-analysis accelerator: a comprehensive tool for statistical data conversion in systematic reviews with meta-analysis. *BMC Med Res Methodol*. 2024;24(1):243. doi:10.1186/s12874-024-02356-6
21. Eticha MD, Hunde AB, Ketema T. Designing a Context-Driven Problem-Solving Method with Metacognitive Scaffolding Experience Intervention for Biology

Instruction. *J Sci Educ Technol*. 2024;33(6):811-822. doi:10.1007/s10956-024-10107-x
22. Kehl J, Krachum Ott P, Schachner M, Civitillo S. Culturally responsive teaching in question: A multiple case study examining the complexity and interplay of teacher practices, beliefs, and microaggressions in Germany. *Teaching and Teacher Education*. 2024;152:104772. doi:10.1016/j.tate.2024.104772
23. McKenney S, Reeves T. *Conducting Educational Design Research*. Taylor & Francis; 2018. doi:https://books.google.co.id/books/about/Conducting_Educational_Design_Research.html?id=aVBtDwAAQBAJ&redir_esc=y
24. McKenney S, Reeves TC. Educational design research for relevant & robust scholarship. *J Comput High Educ*. 2025;37(2):614-638. doi:10.1007/s12528-025-09456-2
25. Azizah N, Inderawati R, Vianty M. Developing descriptive reading materials in EFL classes by utilizing the local culture. *SiELE*. 2021;8(2):596-621. doi:10.24815/siele.v8i2.18562
26. Assaly I, Jabarin A. Arab Israeli EFL teachers' perceptions and practices vis-à-vis teaching higher-order thinking skills: A complicated relationship. *Language Teaching Research*. 2024;28(4):1635-1655. doi:10.1177/13621688211032426
27. Widiastuti I. Membaca Arsitektur Nusantara Mengkini dengan Perspektif Rizomatik Deleuze dan Guattari. *ATRIUM*. 2026;12(1):11-30. doi:10.21460/atrium.v12i1.450
28. Sholikah L, Pertiwi FN. Analysis of Science Literacy Ability of Junior High School Students Based on Programme for International Student Assessement (pisa). *insecta*. 2021;2(1):95-104. doi:10.21154/insecta.v2i1.2922
29. Aldayel HS, Pravitasari H, Fadillah N. Factors affecting academic writing difficulties among English as a Foreign Language (EFL) university students. *Acta Psychologica*. 2026;263:106250. doi:10.1016/j.actpsy.2026.106250
30. Clark-Stallkamp R, Reis N. Design, Development, and Formative Evaluation of a Hybrid Remote Quantum Laboratory Supporting HBCU Partnerships. *J Form Des Learn*. 2024;8(1):1-14. doi:10.1007/s41686-024-00088-w
31. Pijeira-Díaz HJ, Van De Pol J, Channa F, De Bruin A. Scaffolding self-regulated learning from causal-relations texts: Diagramming and self-assessment to improve metacomprehension accuracy? *Metacognition Learning*. 2023;18(3):631-658. doi:10.1007/s11409-023-09343-0
32. De Jong T, Lazonder AW, Chinn CA, et al. Let's talk evidence – The case for combining inquiry-based and direct instruction. *Educational Research Review*. 2023;39:100536. doi:10.1016/j.edurev.2023.100536
33. Siswanto S, Yuliana I. Linking transformational leadership with job satisfaction: the mediating roles of trust and team cohesiveness. *JMD*. 2022;41(2):94-117. doi:10.1108/JMD-09-2020-0293
34. Krause-Wichmann T, Greisel M, Wekerle C, Kollar I, Stark R. Promoting future teachers' evidence-informed reasoning scripts: Effects of different forms of instruction after problem-solving. *Front Educ*. 2023;8:1001523. doi:10.3389/feduc.2023.1001523

35. Habók A, Oo TZ, Magyar A. The effect of reading strategy use on online reading comprehension. *Heliyon*. 2024;10(2):e24281. doi:10.1016/j.heliyon.2024.e24281
36. Li H, Gan Z, Leung SO, An Z. The Impact of Reading Strategy Instruction on Reading Comprehension, Strategy Use, Motivation, and Self-Efficacy in Chinese University EFL Students. *Sage Open*. 2022;12(1):21582440221086659. doi:10.1177/21582440221086659
37. Dekker I, Meeter M. Evidence-based education: Objections and future directions. *Front Educ*. 2022;7:941410. doi:10.3389/feduc.2022.941410
38. UNESCO. *Guidance for Generative AI in Education and Research*. UNESCO; 2023. doi:10.54675/EWZM9535

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

