Arabiyatuna
Jurnal Bahasa Arab

Vol. 10. No. 2, November 2026, 477-502
P-ISSN: 2580-5045, E-ISSN: 2580-5053
DOI: http://dx.doi.org/10.29240/jba.v10i2.17878

Developing and Validating a Strategic Translanguaging Framework for Thematic-Integrative Speaking Skill in Islamic Higher Education

Rojja Pebrian[1*], Saproni[2], Gamal Abdul Nasir[3], Rahmat Iswanto[4], Rahmah Ahmad Osman [5]
Universitas Islam Riau, Indonesia[1,2]
Universiti Brunei Darussalam, Brunei Darussalam[3]
Institut Agama Islam Negeri Curup, Indonesia[4]
International Islamic University Malaysia, Malaysia[5]

[*]Corresponding Author: *rojjapebrian@fis.uir.ac.id*

Cite this article:
Pebrian, R., Saproni, S., Abdul Nasir, G., Iswanto, R., & Ahmad Osman, R. (2026). Developing and Validating a Strategic Translanguaging Framework for Thematic-Integrative Speaking Skill in Islamic Higher Education. Arabiyatuna: Jurnal Bahasa Arab, 10(2), 477–502. https://doi.org/10.29240/jba.v10i2.17878

Received: 31-07-2025	Revised: 21-09-2026	Accepted: 31-10-2026

Abstract

Maharah kalam (speaking skill) is essential for non-Arab learners in Islamic higher education; however, classroom tasks frequently fall short of fostering the extended, theme-bound discourse required in academic contexts. Consequently, Bahasa Indonesia often enters instruction as an unplanned habit rather than a strategically designed and systematically withdrawn scaffold. To address this gap, this study developed and validated a conceptual framework for strategic translanguaging within thematic-integrative *maharah kalam* instruction. The study employed a three-phase design: a PRISMA-adapted literature synthesis of publications from 2021 to 2026 ($n = 28$), a double-coded curriculum analysis within the Arabic Language Education Program at Universitas Islam Riau, and an expert validation using Aiken's V involving five specialists evaluating appropriateness, clarity, and practical utility. The resulting framework integrates input, process, and output components. The instructional process operates through concept explanation, lexical activation, and metacognitive reflection, featuring a structured protocol of translanguaging points and gradual L1 fading. A 100-minute instructional blueprint demonstrates the systematic reduction of L1 reliance within a single unit. Expert evaluation

Copyright © 2026 **Author'(s)**
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

yielded high content validity, with Aiken's V coefficients ranging from 0.900 to 0.933 (V > 0.80) and an overall feasibility score of 3.75 (Very Good). Although clarity initially received the lowest rating and required primary revision, the experts fully endorsed the protocol for scaffolded L1 integration and withdrawal. This paper establishes the pedagogical feasibility of the framework, while acknowledging that empirical measurement of student speaking gains remains a subject for future research.

Keywords: Strategic trans-language; *maharah kalam*; thematic-integrative learning; Arabic language teaching; scaffolded instruction

Introduction

Maharah kalam is one of the skills non-Arab students in Islamic higher education are expected to leave with (Abidah et al., 2026). Classroom practice often falls short of that expectation. Students hear little authentic Arabic, grammar-translation still takes most of the hour, confidence drops once a turn runs longer than a sentence, and the materials sit some distance from the communicative competence written into the curriculum (Rosyad et al., 2026; Abidah et al., 2026). Bibliometric maps of Arabic language education in Indonesian Islamic higher education show a rapidly growing literature. They do not show the pedagogical handling of students' first language as a theme in its own (Maghfirah et al., 2025; Safrudin et al., 2024).

Thematic teaching has become hard to avoid once a cohort is linguistically mixed. Local themes have been used to organize Arabic speaking materials for school students (Irawan et al., 2026) and to select texts in other language classrooms (Gunawardane et al., 2025). Developers of speaking modules have asked whether learners are ready to produce speech, and how far they can work without the teacher (Nurjanah et al., 2021), including in Arabic *maharah kalam* materials (Urwati et al., 2023). Arabic tasks built on Content and Language Integrated Learning (CLIL) have been linked to higher self-efficacy (Kamis et al., 2021). These studies support linking a theme to oral Arabic. However, none of these studies report an L1 protocol for *maharah kalam* in Islamic higher education.

Translanguaging research inside CLIL is fairly consistent on one point. A planned use of the first language helps when the content is cognitively heavy (Hu et al., 2023). The stronger accounts refuse to leave L1 as a rescue device. They treat it as a way of tying what students already know to the material in front of them (Zhou & Mann, 2021; Abourehab & Azaz, 2023). Chinese CLIL classrooms offer a concrete picture: teachers place L1 at specific moments, explain a concept and return to it in reflection, so that students stay in the task

while content and language are both in play (Zhou & Mann, 2021; Hu et al., 2023). Reviews from the same region warn against carrying a European CLIL package into Asian classrooms as if the linguistic setting did not matter. Curriculum design and teacher development have to travel with the method (Hu, 2023). Materials studies are blunter. CLIL texts often pass over what students already know, and they treat cognition, culture, and communication thinly. Classroom use then stalls (Cao, 2021).

A stricter reading is still required. Gains in the target language can look larger than they are when selection and prior preparation go uncontrolled, so the flattering version of CLIL is a poor guide (Feddermann et al., 2021). When materials fail to balance content, cognition, communication, and culture in balance, teachers improvise (Cao, 2021; Hu, 2023). Pedagogical translanguaging has supported learning in Arabic heritage and community classes (Abourehab & Azaz, 2023). Those learners are not the non-Arab undergraduates whose L1 is Indonesian and who hear little Arabic once they leave the room. If L1 is left open-ended, the Arabic practice that does exist shrinks further (Awwaludin et al., 2022; Maulita, 2023). Indonesian *maharah kalam* therefore needs a limited, designed use of L1, with an explicit point at which that support is withdrawn. Studies of Arabic phrase translation point in the same direction: movement between language systems holds when it is procedural, and it frays when it is left to feel (Anis et al., 2022). *Strategic translanguaging*, as used here, means the planned, bounded, and gradually withdrawn use of Bahasa Indonesia at three instructional points—concept explanation, thematic lexical activation, and metacognitive reflection—so that thematic Arabic speech has a scaffold. Incidental code-switching is a different practice. So are teaching the whole session in Indonesian, and immersion that offers no scaffold at all. For Arabic language education, the new step is to treat the function of L1, its timing, and its withdrawal as decisions a lecturer can design inside a thematic *maharah kalam* unit.

Hu and colleagues reserve the word strategic for L1 used at selected moments and aligned with the teaching aim, the difficulty of the content, and who the learners are (Hu et al., 2023; Hu, 2022). In Indonesian classrooms the distance between the task and real talk shows up in three places, and each is a different kind of distance. Speaking work is still often a sentence drill or a topic recitation. Campus and religious talk asks for something longer, and bound to a theme (Rosyad et al., 2026; Abidah et al., 2026). Support, where it exists, usually stops at a grammar point or a word list. A student can recognize the form and still be unable to hold a turn (Rosyad et al., 2026). Assessment in outcome-based Arabic programs records the performance at the end. It rarely records how L1 support is withdrawn (Samin et al., 2026). A similar mismatch, between what the class does and what the profession later asks, has been described in other vocational language programs (Mahmood et al., 2025). The problem is a missing

design: a way to turn Bahasa Indonesia from an unplanned prop into a temporary thematic scaffold, inside a curriculum that already asks for thematic Arabic in real academic situations.

The pressure on that problem comes from several directions at once. Recent innovation agendas in Arabic language learning have gathered around technology—digital media, information and communication technology, newer learning tools, and large language models—while the decision about which language the lecturer uses has stayed in the background (Islami et al., 2026; Muhdi et al., 2024; Shadiev et al., 2023; Raharjo et al., 2026). The bibliometric maps already cited make the same absence visible from the other side. Managing L1 is everyday classroom business, and not a research theme (Maghfirah et al., 2025; Safrudin et al., 2024). Without a fading protocol, Bahasa Indonesia thins an Arabic-speaking environment that is already limited (Awwaludin et al., 2022; Maulita, 2023). An operational framework has to be built and validated before module trials and classroom experiments can mean much.

Earlier studies have tried to thicken the Arabic environment, to build speaking modules or digital media, or to encourage self-regulated learning. An L1 protocol inside the thematic unit is not part of that work (Awwaludin et al., 2022; Maulita, 2023; Urwati et al., 2023; Abidah et al., 2026; Samin et al., 2022). International CLIL research has named critical translanguaging points (Zhou & Mann, 2021; Hu et al., 2023) and has argued for localization through the 4Cs—Content, Cognition, Communication, and Culture (Cao, 2021; Hu, 2023). That argument has not yet been turned into a validated operational framework for thematic-integrative *maharah kalam* in Indonesian Islamic higher education. Three gaps remain. Thematic units still do not mark their translanguaging points. No fading sequence moves students from L1 support toward Arabic they can manage on their own. And such a framework has not been put in front of experts; the literature describes responsive L1 use and stops. What is still missing is a locally adapted, expert-validated protocol that states those points and those procedures for thematic-integrative *maharah kalam*.

The difference from earlier work is concrete. Descriptive accounts of Arabic classrooms treat Bahasa Indonesia as something the teacher reaches for when proficiency runs low (Awwaludin et al., 2022; Maulita, 2023). This study states the moment, the reason, and the later withdrawal. Module and digital-media studies strengthen the speaking task and leave L1 undesigned (Urwati et al., 2023; Abidah et al., 2026; Ruby Syahira & Zukhaira, 2026). Here a bounded translanguaging-point protocol is placed inside the thematic unit. CLIL–translanguaging research has recorded strategic L1 use mainly in content classrooms that are not Arabic ones (Zhou & Mann, 2021; Hu et al., 2023; Li & Zhang, 2022). Those findings are carried into a setting of non-Arab learners in Indonesian Islamic higher education, with the curriculum documents of the Arabic Education Study Program at Universitas Islam Riau as the local base, and

the resulting framework is put through expert content validation. The original step is the conversion of a known insight—that L1 can support learning—into an expert-validated operational framework in which first-language use is a pedagogical scaffold the lecturer designs.

The questions that follow from this gap are these.

1. RQ1: What theoretical foundations and research gaps concerning translanguaging and CLIL are relevant to thematic-integrative *maharah kalam* instruction in Islamic higher education?
2. RQ2: What structure and components should constitute a strategic translanguaging conceptual framework appropriate for the local context?
3. RQ3: To what extent is the proposed framework considered feasible based on expert evaluations of its appropriateness, clarity, and practical usefulness?

The objectives of the study are to identify the theoretical bases and the research gaps concerning strategic translanguaging in thematic-integrative *maharah kalam* for non-Arab learners in Islamic higher education, to build the structure and the main components of a framework adapted from CLIL to this local setting, and to establish whether specialists in Arabic language education, curriculum, and CLIL judge that framework feasible. The expected theoretical contribution is a reconceptualization of Bahasa Indonesia as thematic scaffolding that can be designed, together with a localization of CLIL translanguaging for this kind of class (Hu, 2023; Cao, 2021). The expected practical contribution is a protocol—concept explanation, lexical activation, and metacognitive reflection with fading—that lecturers can insert into *Maharah al-Kalam* units already on the syllabus. At curriculum level, the same protocol supplies a written scaffolding procedure that can sit beside Outcome-Based Education, where the outcomes are explicit and the oral-scaffolding steps are not (Mufanti et al., 2024; Samin et al., 2026). What these objectives are set to establish is expert-validated pedagogical feasibility.

Method

The aim was to develop a strategic translanguaging framework for thematic-integrative *maharah kalam* among non-Arab learners in Islamic higher education. Three streams of evidence were kept in step: a synthesis of recent work on translanguaging within CLIL, an analysis of curriculum documents read against the Indonesian setting, and validation by specialists in Arabic language education, curriculum, and CLIL. The validators judged appropriateness, clarity, and practical usefulness. Findings are reported in the order of the three research questions.

The methodological contribution is that combination. *Maharah kalam* research in Indonesia has mostly taken the form of module trials or descriptive

surveys. CLIL–translanguaging research has mostly taken the form of classroom discourse analysis or systematic review (Zhou & Mann, 2021; Li & Zhang, 2022; Hu et al., 2023; Abidah et al., 2026). Used alone, neither family yields an expert-validated operational protocol for thematic Arabic speaking. Quality criteria from educational design research (Nieveen & Folmer, 2013; Matolić et al., 2023) were adapted to a three-stream design: a PRISMA-adapted conceptual synthesis, double-coded analysis of the local curriculum, and content validation with Aiken's *V*. The product is a content-valid framework that later classroom work can test. The techniques are familiar. Arranged this way, they produce a kind of evidence that *maharah kalam* research in Islamic higher education has not had: pedagogical feasibility, with the translanguaging points specified.

Data Sources and Participants

Three kinds of sources were used. The first was secondary: peer-reviewed articles on translanguaging, CLIL, and *maharah kalam* for non-Arab learners, retrieved from Scopus, Google Scholar, and ERIC. An initial search ran from 5 to 20 January 2026. A further search in June 2026 picked up journal issues that had appeared since, so the synthesis would still be current when the framework was closed. Both rounds used the Boolean string ("translanguaging" OR "first language use" OR "L1 scaffolding") AND ("CLIL" OR "content and language integrated learning") AND ("Arabic speaking" OR "*maharah kalam*" OR "speaking skill"), limited to 2021–2026. Records from the two rounds were pooled and screened against the same criteria. The reporting flow follows PRISMA for the sake of a transparent conceptual synthesis. Effect estimation, and a full systematic-review protocol, were outside the aim. Of 216 records identified, 168 remained after duplicates were removed, and 54 passed title-and-abstract screening. Of those 54, 36 went to full-text assessment. The written screening account does not place the other 18 records in a separate exclusion category, so Figure 1 reports that count without assigning a reason. Of the 36 full texts, 8 were excluded because they did not address translanguaging, CLIL, or speaking pedagogy, or because they were not peer-reviewed journal articles within the 2021–2026 window. Twenty-eight articles remained. Inclusion required a peer-reviewed journal article in English, Indonesian, or Arabic, with Latin transliteration where the Arabic record needed it. Conference papers, books, and grey literature were excluded. Figure 1 shows the funnel.

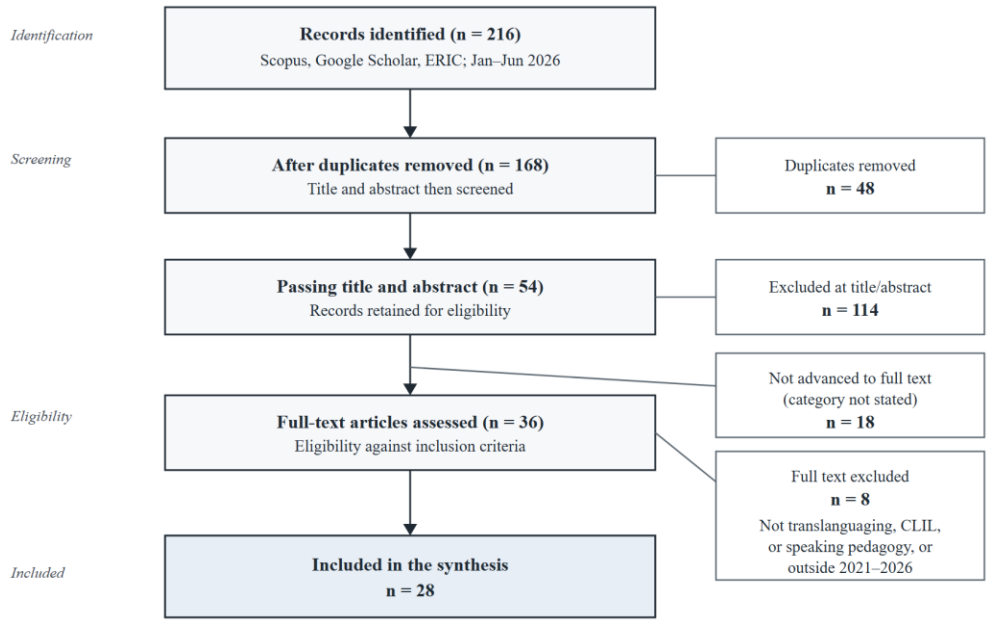


Figure 1. PRISMA-adapted flow of literature identification and screening for the conceptual synthesis.

Note. The procedure is adapted from the PRISMA reporting flow for transparency of conceptual synthesis; it is not claimed as a full systematic review or meta-analysis. Duplicate removal (n = 48) is the difference between records identified (n = 216) and records remaining (n = 168). The remainder (n = 18) is the difference between records passing title-and-abstract screening (n = 54) and full texts assessed (n = 36). The screening account states no exclusion category for those 18 records. The only stated full-text exclusion is the set of 8 records described in the paragraph above.

The second source was documentary: the curriculum of the Arabic Education Study Program, Faculty of Islamic Studies, Universitas Islam Riau. The set comprised the Graduate Learning Outcomes (CPL), the course syllabus, and the Semester Learning Plan (*Rencana Pembelajaran Semester*—RPS) for *Maharah al-Kalam* in 2024/2025. Thematic content analysis was used to locate instructional needs and to mark stages at which first-language (L1) scaffolding would be required. The reading was meant to yield contextual evidence that similar Arabic Education programs in other Islamic higher education institutions could examine for transfer. Units of analysis were the graduate learning outcomes tied to spoken discourse; the thematic units in the syllabus and the RPS; the indicators used to assess speaking; and the instructional strategies named in the documents. Coding moved through three passes. Open coding first marked segments on speaking outcomes, thematic units, L1 scaffolding, assessment, and signs of independent Arabic production. Axial coding then gathered those codes into five categories: L1 functions; translanguaging points; thematic integration; fading mechanisms; and assessment orientation. The

categories were set against widely used CLIL principles to identify the operational gaps and to tighten the framework. The principal researcher and a second researcher coded independently. The two sets were compared segment by segment. When a segment was placed in different categories, the disagreement was recorded and discussed until the two coders assigned it the same way. Credibility here rests on that consensus, not on a single coder.

Primary data came from validation forms completed by five specialists: two in Arabic language pedagogy, one curriculum specialist in Arabic language education, and two in CLIL and translanguaging. They rated the framework on appropriateness, clarity, and practical usefulness.

Research Procedure

The framework was built in a five-stage cycle. Stage 1 drew out the main findings and the instructional patterns of translanguaging in recent CLIL work. Stage 2 mapped what non-Arab learners in Islamic higher education need, using the curriculum documents. Stage 3 asked where strategic translanguaging, thematic-integrative teaching, and the local learners could actually meet. Stage 4 wrote the procedural structure, fixed the critical translanguaging points, and drafted an operational protocol for each thematic unit. Stage 5 sent the framework to the experts and revised it until they converged on feasibility. The stages ran in that order, with room to revise an earlier decision in light of a later one. Stages 1 and 2 answer RQ1, stages 3 and 4 answer RQ2, and stage 5 answers RQ3.

Those five stages are the developmental cycle. Inside the cycle, evidence was handled in four audit steps—collection, classification, analysis, and reporting—following the process logic of recent Arabic-language studies that treat bilingual transfer as a designed procedure (Anis et al., 2022).

Data collection. Three streams were collected together. (1) Secondary literature was retrieved from Scopus, Google Scholar, and ERIC on 5–20 January 2026, with an update in June 2026, using the Boolean string above, limited to 2021–2026. (2) Documentary data comprised the CPL, syllabus, and RPS of the *Maharah al-Kalam* course at Universitas Islam Riau (2024/2025). (3) Primary data comprised completed expert-validation forms from five specialists who met the eligibility criteria set out below.

Data Classification

After duplicates were removed, records were marked included or excluded at title-and-abstract level and again at full text We open-coded curriculum segments into speaking outcomes, thematic units, L1 scaffolding, assessment practices, and indicators of independent Arabic production. Expert ratings were sorted by dimension (appropriateness, clarity, practical usefulness) and by the theme of the comment.

Data Analysis

The 28 articles were synthesized conceptually. They were not meta-analyzed. Curriculum codes were grouped axially into five categories—L1 functions, translanguaging points, thematic integration, fading mechanisms, and assessment orientation—and set against CLIL principles. Expert scores were examined through dimension means and Aiken's *V*. Open comments were reduced to points for revision.

Reporting of results

Findings follow RQ1, then RQ2, then RQ3: the literature-and-curriculum gap (Table 1), the framework and its protocol (Tables 2–4, Figure 2), and the feasibility evidence (Tables 5–6). The sequence—collect, classify, analyse, report—is what makes the conceptual design open to audit and to replication.

Research Instrument

The research instrument consisted of an expert validation questionnaire developed by adapting educational intervention quality criteria—namely relevance (validity), consistency, and practicality—into three evaluation dimensions: (1) the appropriateness of the framework with respect to translanguaging theory and the thematic-integrative approach; (2) the clarity of the presentation of the framework's stages, components, and strategic translanguaging protocol; and (3) the practical usefulness of the framework for maharah kalam instruction in Islamic higher education (Nieveen & Folmer, 2013). The questionnaire comprised 12 closed-ended items, with four items for each evaluation dimension, measured on a four-point Likert scale (1 = Poor, 2 = Fair, 3 = Good, and 4 = Very Good). Each evaluation dimension also included an open-ended section for comments and suggestions. This concise 12-item format is consistent with the development of evaluation instruments for educational and training programs that emphasize practical feasibility without compromising construct coverage (Matolić et al., 2023). The score for each evaluation dimension was calculated as the mean of its four constituent items, whereas each expert's overall score represented the average of the three evaluation dimensions. Before implementation, an expert in Arabic language education research methodology validated the questionnaire for linguistic clarity and content appropriateness. The experts were selected based on the following criteria: (a) a minimum of five years of teaching or research experience in maharah kalam; (b) at least a master's degree in Arabic language education or applied linguistics; and (c) a record of indexed scholarly publications in relevant fields within the previous five years.

Data Analysis

Quantitative ratings were examined with descriptive statistics and a content-validity coefficient. Comments were examined thematically. For each dimension, the mean was read against these bands: 1.00–1.74 = Poor; 1.75–2.49 = Fair; 2.50–3.24 = Good; 3.25–4.00 = Very Good. A mean of 3.00 was the acceptance floor. Means alone do not show whether the raters agreed, so content validity was also estimated with Aiken's V, $V = \Sigma s \,/\, [n(c - 1)]$, where s is the distance between a rater's score and the lowest score on the scale, n is the number of raters ($n = 5$), and c is the number of categories ($c = 4$) (Aiken, 1985). V was computed for each of the 12 items and for each of the three dimensions. From Aiken's critical-value table, $V \geq 0.80$ was required before an item or a dimension could be treated as valid for five raters and four categories ($p < .05$). Comments were reduced, grouped by theme, and turned into revisions. Consensus was defined in advance as a condition in which at least four of the five experts had an individual mean in the Good or Very Good band on all three dimensions. After the ratings came in, the framework was revised, and then closed as the study product.

Data Validation Procedures

Validation was layered to match the three sources. The literature corpus was checked against the inclusion rules, duplicated, and run through the PRISMA-adapted eligibility screen, which left 28 articles (Figure 1). The curriculum analysis was checked by independent double coding; disagreements were recorded and resolved in discussion until the coders agreed on every disputed segment. The framework itself was rated by five experts on appropriateness, clarity, and practical usefulness. Dimension means were read against an acceptance floor of 3.00, and Aiken's V against a critical value of 0.80 ($n = 5$, $c = 4$, $p < .05$). Comments were analyzed thematically and used to revise the protocol before the framework was closed. The procedure supports a claim of content-valid feasibility. It does not support a claim of classroom effectiveness.

Ethical Considerations

All five validators gave informed consent. Identities were replaced with the codes E1–E5, and any expert could withdraw at any time without penalty. Only expert review was involved. No participants were vulnerable, and the work was treated as minimal-risk research. Procedures followed the research ethics guidelines of Universitas Islam Riau.

Findings and Discussion

Theoretical Foundations and Literature Gap Analysis (RQ1)

RQ1 is answered by placing international CLIL practice alongside *maharah kalam* instruction in Indonesian Islamic higher education, and alongside the research site's curriculum documents. International studies put strategic first-language (L1) use at points decided in advance—concept explanation, clarification, and metacognitive reflection—and they treat teacher scaffolding as mediation that loosens, from linguistic support toward production the student manages (Hu et al., 2023; Hu, 2022; Zhou & Mann, 2021; Li & Zhang, 2022). Where that design is taken up by the institution, curriculum alignment and teacher development become what Luo and Chen (2022) describe as a measurable side of instructional capability (Hu, 2023). Indonesian practice still reaches for Bahasa Indonesia when difficulties arise. Grammar-centered materials dominate, the Arabic-speaking environment is thin, and L1 fills a proficiency gap instead of scaffolding thematic speech (Rosyad et al., 2026; Awwaludin et al., 2022; Maulita, 2023). CLIL materials have their own imbalance. Content, cognition, communication, and culture are often left uneven, and the teachers improvise (Cao, 2021; Hu, 2023). Work Arabic language environments, speaking modules, and on self-regulated learning has likewise stalled before a validated protocol for translanguaging points and fading (Awwaludin et al., 2022; Maulita, 2023; Urwati et al., 2023; Abidah et al., 2026; Samin et al., 2022). Table 1 puts the contrast in one place.

Table 1. Synthesis of the Literature Gap between International CLIL Research and *Maharah Kalam* Practices in Indonesian Islamic Higher Education

No	Comparative Aspect	International CLIL Literature	***Maharah Kalam*** **Practices in Islamic Higher Education**
1	Function of L1	Structured thematic scaffolding	Responsive and intuitive instructional medium
2	Translanguaging points	Explicitly defined at critical instructional stages	No explicit operational protocol
3	Thematic integration	The 4Cs are a design aim, yet materials often leave content, cognition, communication, and culture unbalanced	Thematic units are listed, but they are not linked to translanguaging points
4	Fading mechanism	Gradual withdrawal of L1 toward autonomous production, with L1 retained only at justified points	Not formulated; L1 is neither timed nor withdrawn
5	Institutional support	Supported by curriculum policies and teacher training	Not yet standardized within the curriculum

Note. L1 = first language. The 4Cs = Content, Cognition, Communication, and Culture. The international CLIL column synthesizes Hu et al. (2023), Zhou and Mann (2021), Cao (2021), Hu (2023), Li and Zhang (2022), and Luo and Chen (2022). Row 3 records the critical finding that CLIL materials often remain unbalanced (Cao, 2021; Hu, 2023). Row 4 records graduated scaffolding (Li & Zhang, 2022; Zhou & Mann, 2021) and matches the fading definition used in this study: Bahasa Indonesia is withdrawn until Arabic is the primary medium, while L1 stays optional at designated points. The *maharah kalam* column synthesizes Rosyad et al. (2026), Abidah et al. (2026), Awwaludin et al. (2022), Maulita (2023), and the curriculum document analysis conducted in this study.

Table 1 is a contrast the procedures. The fading used in this framework is the one in row 4: Arabic becomes the main medium, and Bahasa Indonesia stays available only at points that have been marked (Li & Zhang, 2022; Zhou & Mann, 2021). Reviewing the Graduate Learning Outcomes (CPL), the syllabus, and the Semester Learning Plan (RPS) of the Arabic Education Study Program at Universitas Islam Riau revealed three patterns. Intended outcomes ask for communicative Arabic discourse and rarely say how L1 is to scaffold it. Thematic units are arranged by topic, with no translanguaging point for concept explanation, lexical activation, or reflection. Speaking assessment stores the final performance. Nothing in it withdraws L1 support by stages. The same split—outcomes and assessment written down, oral scaffolding left thin—has been reported for Outcome-Based Education more broadly (Samin et al., 2026; Mufanti et al., 2024). Those three patterns became the input side of the framework. The RQ1 result is an operational gap. Translanguaging points, fading, and expert-validated procedures for thematic-integrative *maharah kalam* are not yet specified in this setting.

Structure of the Strategic Translanguaging Conceptual Framework (RQ2)

The framework answering RQ2 is an input–process–output sequence. Input takes in learner needs, the mapping of themes, and materials drawn from CLIL literature and from teaching practice in Islamic higher education. Process is the three translanguaging phases below. Output sets thematic Arabic speaking, critical thinking (Anwar & Hilmi, 2025), and readiness for real academic and non-academic talk as design targets. Each stage has to be in place before the next one can do its work.

Tables 2 and 3 list the components and the protocol points. The rule that orders the three process phases is simple enough to apply. In Phase 1 the lecturer uses Bahasa Indonesia to introduce the theme and the aims of the session. Therefore, a new concept is settled before anyone is asked to speak Arabic. In Phase 2 students pair Arabic terms with Indonesian ones—bilingual mapping, vocabulary matching, short dialogues—along a scaffold that moves from linguistic support toward production they can manage (Li & Zhang, 2022). In Phase 3, reflection and oral production are in Arabic. Bahasa Indonesia comes back only when a point will not clarify in Arabic. Before the session, the lecturer marks where L1 is allowed: a new concept, an otherwise unclear task

instruction, genuinely complex thematic vocabulary.. L1 stays inside that list. Figure 2 shows how input, process, and output sit inside one thematic unit. Concept explanation belongs to process. Input is finished before the session starts.

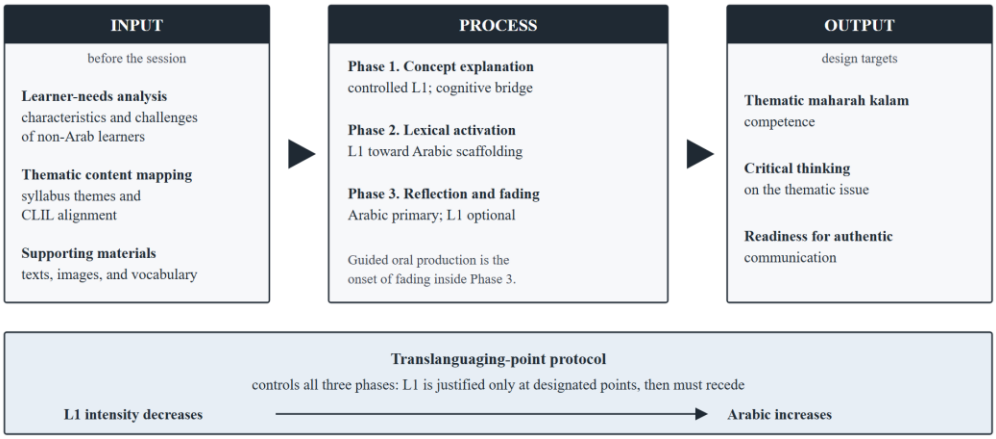


Figure 2. Conceptual framework of strategic translanguaging for thematic-integrative *maharah kalam* instruction.

Note. L1 = first language (Bahasa Indonesia); TL = target language (Arabic). The diagram is read from left to right. The Input dimension comprises learner-needs analysis, thematic content mapping, and supporting instructional materials, and is completed before the session. The Process dimension comprises three sequential translanguaging phases, namely concept explanation, lexical activation, and metacognitive reflection with fading. The Output dimension comprises thematic speaking competence, critical thinking, and readiness for authentic communication; these are design targets, not measured learning gains. The translanguaging-point protocol functions as the controlling mechanism operating across all three phases within each thematic unit. The protocol specifies where L1 is pedagogically justified and where it must recede.

Table 2. Dimensions and Components of the Strategic Translanguaging Conceptual Framework for Thematic-Integrative *Maharah Kalam* Instruction

No	Dimension	Core Component	Functional Description
1	Input	Learner Needs Analysis	Analysis of the characteristics, needs, and challenges of non-Arab learners in Islamic higher education to inform the design of thematic and translanguaging-based *maharah kalam* instruction.
2	Input	Thematic Content Mapping	Identification of relevant themes or instructional units based on the course syllabus and classroom needs, integrating findings from CLIL literature and instructional practices in Islamic higher education.

3	Input	Supporting Instructional Materials	A collection of instructional resources, including texts, images, and videos, designed to facilitate thematic understanding and the acquisition of essential vocabulary.
4	Process	L1-Based Thematic Concept Explanation	Controlled use of learners' first language (L1) to explain new concepts, establish cognitive bridges, and minimize misconceptions.
5	Process	Thematic Lexical Activation through Scaffolding	Development of thematic vocabulary through translanguaging, collaborative discussion, brain mapping, and theme-based dialogue, gradually progressing from L1 support to Arabic language production.
6	Process	Metacognitive Reflection and Translanguaging Fading	Activities involving the reconstruction, reflection, and evaluation of conceptual understanding and language-learning strategies, accompanied by the gradual fading of L1 support toward independent Arabic language production.
7	Output	Improved Thematic *Maharah Kalam* Competence	Learners demonstrate the ability to communicate confidently in Arabic on thematic topics within authentic communicative contexts.
8	Output	Enhanced Critical Thinking	Learners demonstrate the ability to analyze, synthesize, and critically reflect on thematic issues through Arabic.
9	Output	Readiness for Authentic Communication	Learners are prepared to communicate effectively in real-life academic and non-academic situations, with progressively reduced reliance on translanguaging and greater independence in Arabic language use.

Note. L1 = first language. Rows 7–9 state the design targets of the framework. They are not measured learning outcomes from a classroom trial.

Table 3. Operational Protocol for Strategic Translanguaging Points Across Thematic Instructional Phases

No	Phase	Strategic Translanguaging Point	Pedagogical Purpose	Sample Activities and Implementation
1	Phase 1	Thematic concept	To facilitate	The lecturer explains

	(Process)	explanation at the beginning of the instructional session	learners' understanding of thematic concepts and activate their prior knowledge.	new thematic terms or concepts using controlled L1 support. Learning objectives are clarified in L1 to ensure that students fully understand the direction and expectations of the instructional activities.
2	Phase 2 (Process)	Thematic lexical activation during collaborative learning and discussion	To develop thematic vocabulary and establish lexical associations between L1 and Arabic as a form of instructional scaffolding.	Students participate in group discussions using bilingual (L1–Arabic) brain mapping, vocabulary matching activities, and short theme-based dialogues that progressively transition toward Arabic as the primary language of communication.
3	Phase 3 (Process)	Metacognitive reflection and translanguaging fading at the end of the session or after task completion	To evaluate and reflect upon learners' language-learning strategies and learning processes while gradually reducing L1 support in favor of independent Arabic language production.	Students engage in reflective discussions and respond to analytical questions primarily in Arabic, whereas L1 is used only optionally when clarification is genuinely required.
4	Output	Independent thematic *maharah kalam* production with minimal L1 use	To consolidate the design target of independent Arabic use within the thematic unit.	Students deliver thematic presentations or conduct microteaching sessions entirely in Arabic without L1 intervention, except when clarification is absolutely necessary.

Note. The input dimension (learner-needs analysis, thematic mapping, and materials) is specified in Table 2 and is completed before the session. Rows 1–3 are the three process phases. Row 4 is the output check within the same thematic unit, not a fourth translanguaging phase.

RQ2, in short, binds three dimensions, three translanguaging phases, and a translanguaging-point protocol for every thematic unit. The result is a

guide a lecturer can use when planning a session. It is not a description of L1 used on the spur of the moment. Table 4 makes the protocol easier to picture: one 100-minute thematic session of 100 minutes, under the title titled "Academic Discussion on Campus."The table is a design sketch. It is not a report of a class that was taught, and it is not an empirical finding.

Table 4. Illustrative Implementation of the Strategic Translanguaging Protocol in a Single Thematic Instructional Session

No	Time Allocation	Instructional Phase	Strategic Translanguaging Point (L1/Arabic)	Student Activities	Lecturer Observation Indicators
1	0–15 min	Phase 1: Concept explanation	Controlled L1 predominates; key terminology is introduced bilingually.	Students record bilingual (L1–Arabic) equivalents of key terminology after the lecturer introduces the objectives and the theme "Academic Discussion."	Students are able to restate the instructional objectives in L1 and correctly identify at least five key terms in Arabic.
2	15–45 min	Phase 2: Lexical activation	L1-to-Arabic scaffolding; short dialogues progressively transition to Arabic.	Students engage in bilingual brain mapping, vocabulary matching, and two- to three-turn dialogue practice with gradual fading of L1 support.	Students use thematic Arabic vocabulary appropriately in short dialogues while demonstrating progressively reduced reliance on L1.
3	45–75 min	Phase 3 onset: Guided oral production	Arabic serves as the primary language; L1 is used only for essential clarification.	Students participate in thematic group discussions followed by a two-minute individual micro-speaking activity.	Students sustain oral communication in Arabic with only minimal requests for clarification in L1.
4	75–100 min	Phase 3: Reflection and output	Arabic is the primary language; L1 is used optionally during metacognitive reflection.	Students deliver mini-presentations, provide peer feedback, and reflect on their language-	Students complete their mini-presentations in Arabic and identify one personal strategy for improving

		learning strategies. The lecturer identifies fading points for subsequent instructional sessions.	their future speaking performance.

Note. L1 = first language (Bahasa Indonesia). Rows 3 and 4 together constitute Phase 3. Guided oral production is the onset of fading within that phase, not a fourth framework dimension. The proportion of L1 use is expected to decline from row 1 to row 4 within a single session, and across successive units throughout the semester. Observation indicators are design checkpoints for lecturers, not measured learning outcomes.

Expert Validation Results and Framework Feasibility Assessment (RQ3)

RQ3 reports how the framework was presented to experts, and what they thought about its appropriateness, clarity, and practical usefulness. Five experts were recruited on the criteria given in the Method. Two work in Arabic language pedagogy, one in the Arabic language education curriculum, and two in CLIL and translanguaging. All five are senior academics, curriculum developers, or Arabic language education practitioners in Indonesian Islamic higher education. Four hold doctorates and one a master's degree, with 8 to 18 years of professional experience. For confidentiality they are coded E1–E5.

Validation was done online. Each expert received a documentation pack: the framework (input–process–output, the three phases, and the protocol of translanguaging points across thematic units) and the questionnaire. Ratings used a four-point scale (1 = Poor, 2 = Fair, 3 = Good, 4 = Very Good) on appropriateness, clarity, and practical usefulness, with space for open comment. The window was three weeks in March 2026. Each expert had 10 working days. Once the forms were in, dimension means were calculated, Aiken's V was computed, and the comments were read thematically. Table 5 summarises the coefficients. Table 6 gives each expert's scores.

Table 5. Summary of Expert Validation Results and Content Validity Coefficients

No.	Evaluation Dimension	Mean Score	Category	Aiken's *V*	Validity Decision
1	Appropriateness	3.80	Very Good	0.933	Valid
2	Clarity	3.70	Very Good	0.900	Valid
3	Practical Usefulness	3.75	Very Good	0.917	Valid
4	Overall	3.75	Very Good	0.917	Valid

Note. n = 5 raters; c = 4 response categories; each dimension comprises four items. Aiken's V was computed as $V = \Sigma s / [n(c - 1)]$. The critical value for n = 5 and c = 4 at $p < .05$ is 0.80 (Aiken, 1985); coefficients at or above this threshold indicate acceptable content validity.

Table 6. Individual Expert Validation Scores (E1–E5)

Expert Code	Appropriateness	Clarity	Practical Usefulness	Individual Mean
E1	4.00	4.00	4.00	4.00
E2	4.00	3.75	4.00	3.92
E3	4.00	3.75	3.75	3.83
E4	4.00	3.75	3.75	3.83
E5	3.00	3.25	3.25	3.17
Mean	3.80	3.70	3.75	3.75

Note. Each cell represents the mean of four items rated on a four-point scale; consequently, all values are multiples of 0.25. Individual means are rounded to two decimal places.

Tables 5 and 6 place the framework high on all three dimensions. The overall mean is 3.75, inside Very Good. Aiken's V supports that picture: 0.900 to 0.933 at the dimension level, and 0.867 to 1.000 at the item level. Every coefficient exceeds 0.80, the critical value for five raters and four categories, so the agreement is not an artifact of averaging. The consensus rule is also met. All five individual means fall in Good or Very Good—E1 to E4 in Very Good, E5 in Good—above the floor of four out of five set in the Method.

Clarity is the awkward result. It remains Very Good (mean = 3.70; V = 0.900), and still clears the acceptance line, but it sits below appropriateness (3.80; V = 0.933) and practical usefulness (3.75; V = 0.917). Some experts wanted the protocol to be harder to misread: a concrete task at each translanguaging point, and a fading sequence described tightly enough to teach from. Table 6 shows where that pull came from. E2, E3, and E4 gave clarity 3.75, below their appropriateness scores. E5 gave clarity 3.25. The dimension mean dipped, but remained in Very Good and above the validity threshold.

The comments were specific. E2 asked for "concrete task examples at each strategic translanguaging point so that lecturers do not interpret the protocol differently." E5 wrote that "the fading mechanism should be described through operational procedures for each instructional session rather than merely as a general principle." E3 stressed "providing brief training for lecturers before full implementation so that the use of L1 does not extend beyond the designated translanguaging points." The framework was revised against those Three remarks: the illustrative activities were filled out completed, including Table 4 among them; the technical notes on implementation were expanded; and the fading sequence inside in the protocol was tightened before closing the framework. RQ3 therefore records a feasible framework (overall mean = 3.75;

Aiken's $V = 0.917$; consensus from five of five experts). The revision was aimed at operational clarity. It was not offered as evidence that the framework changes what students do in class.

Discussion and Implications of the Proposed Framework

The findings sit in an uneasy relation to the CLIL literature, and to recent Indonesian work on Arabic language education. International studies argue that a deliberate use of the learners' first language can make hard content easier to grasp and can still feed growth in the target language, so long as L1 keeps the job of scaffolding and does not take the lesson over (Hu et al., 2023; Zhou & Mann, 2021). They also want CLIL housed in a wider arrangement: content aims, language aims, and the professional development of teachers (Hu, 2023; Luo & Chen, 2022). Indonesian Arabic classrooms have not settled the argument. One line of work worries that heavy L1 use eats into an Arabic-speaking environment that is already thin (Awwaludin et al., 2022; Maulita, 2023). Another asks for teaching that can bend toward the mixed profiles of Indonesian university students (Rosyad et al., 2026; Samin et al., 2022). Materials are part of the trouble. Too often they miss the students' cognitive starting point, and the 4Cs are weakly joined, which is why an operational translanguaging framework is a more pressing task than importing another ready-made package (Cao, 2021). CLIL also has to notice who the students are. In Norway, reasons for choosing a CLIL program differed by socio-economic setting. A design cannot be assumed to travel unchanged (Bakken & Brevik, 2023). The framework here takes a narrower path. Translanguaging is a planned intervention at a few instructional stages. That stays with thematic-integrative CLIL, and it answers to the conditions of Arabic teaching in Indonesia.

Set against three neighboring literatures, the contribution is easier to locate. Studies of Arabic language environments treat L1 as background, or as a problem to be reduced, and they do not write an operational protocol (Awwaludin et al., 2022; Maulita, 2023). Module and media studies improve the speaking task and do not design L1 as a scaffold (Urwati et al., 2023; Abidah et al., 2026; Ruby Syahira & Zukhaira, 2026). CLIL–translanguaging research records strategic L1 use mainly where the content classroom is not an Arabic one (Zhou & Mann, 2021; Hu et al., 2023; Li & Zhang, 2022). Those separate insights are here turned into one content-valid protocol, built for this setting. The variable—*strategic translanguaging*—is defined by what L1 is for, when it appears, and when it is withdrawn. Hearing Bahasa Indonesia in the room is not, by itself, the variable.

The theoretical stake is a change of status. Indonesian, the students' L1, is treated as a cognitive scaffold under rules, in the service of Arabic oral proficiency. The expert ratings indicate that L1, used at the points already named—concept explanation, thematic lexical activation, metacognitive

reflection—can support comprehension and readiness to speak, while the route toward Arabic that student manages alone stays open. The position keeps CLIL and retunes it for non-native Arabic learners in Indonesia. The claim stops at the quality of the design. Measured learning gain is a different claim, and this study does not make it. In this sense, translanguaging is a pedagogical strategy that connects theory to the classroom. Whether it works once the session is actually taught remains a question for classroom research.

For lecturers who teach *maharah kalam* without a structured set of resources, this is the part they can take into class (Rosyad et al., 2026; Cao, 2021). It sequences a thematic-integrative session, from explaning of the concept to oral Arabic on that theme, and it lets the lecturer decide where Bahasa Indonesia is allowed without cutting the Arabic short. The protocol can also be read against holistic assessment rubrics that want both process and product to leave a trace (Samin et al., 2026). None of this runs by itself. Lecturers have to be ready to mediate the change. CLIL research keeps finding that course management, content knowledge, and lived experience of thematic teaching matter more than another stack of materials (Raharjo et al., 2026; Hu, 2023; Luo & Chen, 2022). Room must also be left for different motives and basic psychological needs (Shelton-Strong, 2022), and for sociocultural background, or the framework collapses into one recipe (Bakken & Brevik, 2023). At the curriculum level it can inform revision as a practical extension of Outcome-Based Education, student-centered learning, and holistic assessment in Arabic Language Education. Therefore, the design meets non-native learners where they are (Pebrian & Samin, 2026; Raharjo et al., 2026; Mufanti et al., 2024). What the validation adds to the contribution is limited, and it is real: the framework is conceptually coherent, content-valid, and fit to be developed further for thematic-integrative *maharah kalam* in Islamic higher education. The experts highlighted the integration of the thematic phases, and the order of the translanguaging process, as strengths. They also asked for tighter technical guidance. RQ1, RQ2, and RQ3 are one argument. The first names an operational gap, the second offers a framework for it, and the third records that the experts found that framework feasible.

Limitations and Directions for Future Research

Four limits should stay in view. The framework has been built, and it has passed expert review. It has not been taught in a *maharah kalam* class, so any talk of effectiveness is talk of feasibility—theoretical and practical—and not of a demonstrated effect on learning. That caution matches evidence that CLIL effects on foreign-language competence can be overstated when selection and prior preparation are left uncontrolled (Feddermann et al., 2021). The synthesis covers only peer-reviewed journal articles from 2021 to 2026 in Scopus, Google Scholar, and ERIC. Grey literature, and anything outside that window, is simply absent. Five specialists were enough for Aiken's *V*, and the coefficients cleared

the critical value, but a panel this small leaves the estimates rough. Non-native Arabic learners, and lecturers from institutions other than the research site, were not among the raters. Five raters and four categories make $n = 5$ admissible for Aiken's V. The coefficients remain evidence of content-valid feasibility in this panel. They do not estimate how lecturers across Islamic higher education would receive the framework. The curriculum reading comes from one Arabic Language Education program at a private Islamic university. What the documents imply for other kinds of Islamic higher education is a further question, and it needs comparison across institutions.

Those limits point to the next steps: materials built on the translanguaging protocol, module evaluation across institutions, notes a lecturer can teach from, and work on pedagogical readiness (Raharjo et al., 2026; Abidah et al., 2026). Later studies can ask what sustained use does to speaking performance, to achievement, and to motivation. Until that work is done, the framework is a conceptually validated base. It is ready for research and development, and for an empirical test of *maharah kalam* modules. It is not evidence that the instruction already works.

Conclusion

Three answers follow the three questions. In thematic *maharah kalam* for non-Arab learners, the core problem is a missing procedure: translanguaging points and a fading mechanism are still unspecified (RQ1). These procedures can be arranged as an input–process–output framework whose process includes concept explanation, lexical activation, and metacognitive reflection, with Bahasa Indonesia withdrawn by stages (RQ2). Five experts judged the framework feasible (overall mean = 3.75, Very Good; Aiken's V = 0.900–0.933). Clarity was the dimension they most wanted tightened (RQ3).

The original contribution is a content-valid pedagogical protocol that treats the learners' first language as a component of thematic speaking instruction in Islamic higher education which lecturers can design. No improvement in speaking scores is reported, and the protocol is not claimed to fit every institutional type without further adaptation. The immediate use is practical. Lecturers and curriculum designers can plan where Bahasa Indonesia enters an existing *Maharah al-Kalam* unit and where it leaves. The next step is a classroom trial and replication in more than one institution, through research and development. On that basis the framework can be turned into translanguaging materials, module trials, and lecturer-preparation programs in Islamic higher education.

Acknowledgment

We thank the Arabic Language Education Program, Faculty of Islamic Studies, Universitas Islam Riau, for access to the curriculum documents—the Graduate Learning Outcomes (CPL), the syllabus, and the *Maharah al-Kalam* Semester Learning Plan (RPS)—which were the basis of the curriculum analysis. We also thank the five expert validators (E1–E5) for their evaluations and for comments that changed the protocol, and the colleagues who cross-checked the document coding.

Author Contributions Statement

RP conceived the framework, led the literature synthesis, and drafted the manuscript. S analized the curriculum document and tightened the operational protocol. GN built the expert validation instrument and interpreted the scores. RI read the manuscript against scholarship in Arabic language education and sharpened the argument for the Indonesian setting. RO critically reviewed the CLIL/translanguaging framework and edited the manuscript. All authors approved the final version.

References

Abidah, A. L., Setiyawan, A., Fuady S, M. N. A. Q., & Putri, N. A. A. (2026). Teaching *maharah al-kalam* through digital media in higher education. *Arabiyatuna: Jurnal Bahasa Arab, 10*(1), 187–206. https://doi.org/10.29240/jba.v10i1.13553

Abourehab, Y., & Azaz, M. (2023). Pedagogical translanguaging in community/heritage Arabic language learning. *Journal of Multilingual and Multicultural Development, 44*(5), 398–411. https://doi.org/10.1080/01434632.2020.1826496

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement, 45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Anis, M. Y., Malik, A., & Setiyadi, A. C. (2022). Developing creative writing process based on the translation of Arabic phrases: Case study in Arabic prose and poetry. *Journal of Narrative and Language Studies, 10*(20), 173–187. https://doi.org/10.59045/nalans.2023.3

Anwar, M. R., & Hilmi, D. (2025). Students' critical thinking skills and Arabic learning outcomes using Quizlet media. *Ijaz Arabi Journal of Arabic Learning, 8*(1). https://doi.org/10.18860/ijazarabi.v8i1.30206

Awwaludin, M., Malik, S., & Siswanto, N. D. (2022). Pembentukan lingkungan bahasa Arab dalam meningkatkan penguasaan bahasa Arab pada

pesantren bahasa Arab (MIM LAM). *Definisi: Jurnal Agama dan Sosial-Humaniora, 1*(1), 55–64. https://journal.uinsgd.ac.id/index.php/definisi/article/view/16716

Bakken, J., & Brevik, L. M. (2023). Challenging the notion of CLIL elitism: A study of secondary school students' motivation for choosing CLIL in Norway. *TESOL Quarterly, 57*(4), 1091–1114. https://doi.org/10.1002/tesq.3173

Cao, Y. (2021). Exploring content and language integrated learning (CLIL) teaching materials in Chinese universities: Teachers' and students' perceptions and reflections. *International Journal of English and Literature, 12*(3), 23–32. https://doi.org/10.5897/IJEL2021.1444

Feddermann, M., Möller, J., & Baumert, J. (2021). Effects of CLIL on second language learning: Disentangling selection, preparation, and CLIL-effects. *Learning and Instruction, 74*, 101459. https://doi.org/10.1016/j.learninstruc.2021.101459

Gunawardane, R., Matthews, J., & Veliz, L. (2025). Identifying themes of curiosity for reading materials in the Sri Lankan ESL university context. *Asian-Pacific Journal of Second and Foreign Language Education, 10*, 49. https://doi.org/10.1186/s40862-025-00357-6

Hu, H. (2022). Factors pertinent to first language use in foreign language classroom: A case of content and language integrated learning. *Arab World English Journal, 13*(2), 177–191. https://doi.org/10.24093/awej/vol13no2.12

Hu, H. (2023). Teaching language and content in multicultural and multilingual classrooms: CLIL and EMI approaches. *LEARN Journal: Language Education and Acquisition Research Network, 16*(1), 753–758. https://so04.tci-thaijo.org/index.php/LEARN/article/view/263462

Hu, H., Said, N. E. M., & Hashim, H. (2023). Sustaining content and language integrated learning in China: A systematic review. *Sustainability, 15*(5), 3894. https://doi.org/10.3390/su15053894

Irawan, B., Yasmar, R., Salsabila, A., Ulum, F., & Turki, M. (2026). Development of a local-culture-based Arabic conversation book to support junior high school students' *mahārah kalām*. *Arabiyatuna: Jurnal Bahasa Arab, 10*(1), 267–290. https://doi.org/10.29240/jba.v10i1.16207

Islami, A., Ibrohim, B., Sudarmawan, H. M., Muttaqien, A., & Lidinillah, F. M. (2026). Technological transformations in Arabic language learning: A comprehensive review of emerging tools and techniques. *Arabiyatuna: Jurnal Bahasa Arab, 10*(1), 227–244. https://doi.org/10.29240/jba.v10i1.15109

Kamis, M. S., Ismail, M. J., Alias, M. N., Mikeng, D., Abidin, S. G. Z., & Yusof, R. (2021). CLIL approach in encouraging self-efficacy amongst Malaysian gifted students for Arabic tasks accomplishment. *Journal of Language and Linguistic Studies, 17*(2), 1001–1012. https://doi.org/10.52462/jlls.69

Li, D., & Zhang, L. (2022). Exploring teacher scaffolding in a CLIL-framed EFL intensive reading class: A classroom discourse analysis approach. *Language Teaching Research, 26*(3), 333–360. https://doi.org/10.1177/1362168820903340

Luo, W.-H., & Chen, Y.-C. (2022). Constructing a teaching capability maturity model for content and language integrated learning teachers in Taiwan. *Humanities and Social Sciences Communications, 9*, 8. https://doi.org/10.1057/s41599-021-00928-1

Maghfirah, T., Himayah, & Hamka. (2025). A decade of Arabic language education research in Indonesian Islamic higher education: A bibliometric mapping of theses and dissertations (2015–2024). *ELOQUENCE: Journal of Foreign Language, 4*(3), 683–704. https://ejournal.iaingorontalo.ac.id/index.php/ELOQUENCE/article/view/3022

Mahmood, A. S., Saad, N. S. M., & Nur, N. M. (2025). Taking flight with Aviation English: How are Malaysian teachers preparing the next generation? *F1000Research, 14*, 1036. https://doi.org/10.12688/f1000research.169220.1

Matolić, T., Jurakić, D., Greblo Jurakić, Z., Maršić, T., & Pedišić, Ž. (2023). Development and validation of the EDUcational Course Assessment TOOLkit (EDUCATOOL) – a 12-item questionnaire for evaluation of training and learning programmes. *Frontiers in Education, 8*, 1314584. https://doi.org/10.3389/feduc.2023.1314584

Maulita, A. A. N. A. (2023). Problematika lingkungan bahasa Arab di Pondok Pesantren Raudlatul Malikiyah Ketapang Kota Probolinggo. *Lisanuna: Jurnal Ilmu Bahasa Arab dan Pembelajarannya, 13*(1), 1–10. https://doi.org/10.22373/ls.v13i1.18025

Mufanti, R., Carter, D., & England, N. (2024). Outcomes-based education in Indonesian higher education: Reporting on the understanding, challenges, and support available to teachers. *Social Sciences & Humanities Open, 9*, 100873. https://doi.org/10.1016/j.ssaho.2024.100873

Muhdi, A., Yuslam, Y., Ngarifillaili, N., Iqbal, A., & Purnama, Y. (2024). Use of ICT in Arabic language education at high school level and its

relationship with teacher skills. *Journal Neosantara Hybrid Learning, 2*(1), 284–301. https://doi.org/10.55849/jnhl.v2i1.746

Nieveen, N., & Folmer, E. (2013). Formative evaluation in educational design research. In T. Plomp & N. Nieveen (Eds.), *Educational design research* (pp. 152–169). Netherlands Institute for Curriculum Development (SLO).

Nurjanah, R. L., Saptanto, D. D., & Dewi, M. K. (2021). Developing modelling-based speaking module for informal interaction to support independent learning of speaking skill. *ELT-Lectura: Studies and Perspectives in English Language Teaching, 8*(1), 99–111. https://doi.org/10.31849/elt-lectura.v8i1.5660

Pebrian, R., & Samin, S. M. (2026). Student centered learning in Arabic language education: Integrating research, community engagement, and microteaching. *Alsinatuna: Journal of Arabic Linguistics and Education, 11*(2), 155–169. https://e-journal.uingusdur.ac.id/index.php/alsinatuna

Raharjo, H., Nurjanah, N., Akbar, G., Mulyadi, Ruswandi, U., Erihadiana, M., & Rahmi, A. (2026). Readiness of Arabic language teachers to integrate large language models (LLMs) in their teaching practices: Challenges and opportunities. *Arabiyatuna: Jurnal Bahasa Arab, 10*(1), 207–226. https://doi.org/10.29240/jba.v10i1.15202

Rosyad, M. S., Rosyadi, F. I., Huda, N., Paramita, N. P., & Guangli, M. (2026). From grammar-centered to communicative materials: Rethinking Arabic language teaching resources in higher education. *Arabiyatuna: Jurnal Bahasa Arab, 10*(1), 103–122. https://doi.org/10.29240/jba.v10i1.16227

Ruby Syahira, & Zukhaira. (2026). Development of Kalam Flipbook Media for Arabic speaking skills of students Grade X Madrasah Aliyah in Semarang City. *Arabiyatuna: Jurnal Bahasa Arab, 10*(1), 123–144. https://doi.org/10.29240/jba.v10i1.15817

Safrudin, R., Nandang, A., Siregar, S. D. P., Musthafa, I., Fauzi, M. F., Alby, M. H. F., & Suyono, S. (2024). Development of Arabic language learning research: A bibliometric study on Scopus (2009-2024). *Al-Ta'rib: Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya, 12*(2), 321–338. https://doi.org/10.23971/altarib.v12i2.8929

Samin, S. M., Akzam, I., & Supriady, H. (2022). Strategies of Arabic students' self-regulated learning improvement in language proficiency in the disruption era. *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban, 9*(2), 165–177. https://doi.org/10.15408/a.v9i2.22828

Samin, S. M., Jaafar, A., Supriady, H., Pebrian, R., Zulkifli, A., & Yunita, Y. (2026). Measuring what matters: Goal-free evaluation of holistic

assessment in Arabic language education. *Ijaz Arabi Journal of Arabic Learning, 9*(1), 278–290. https://doi.org/10.18860/ijazarabi.v9i1.36779

Shadiev, R., Wen, Y., Uosaki, N., & Song, Y. (2023). Future language learning with emerging technologies. *Journal of Computers in Education, 10*(3), 463–467. https://doi.org/10.1007/s40692-023-00285-9

Shelton-Strong, S. J. (2022). Advising in language learning and the support of learners' basic psychological needs: A self-determination theory perspective. *Language Teaching Research, 26*(5), 963–985. https://doi.org/10.1177/1362168820912355

Urwati, N. S., Ruwaida, & Suaidi. (2023). Tathwir muqarrar maharah al-kalam li al-fasl al-'asyir fi al-madrasah al-'aliyah mu'allimat Nahdhat al-Wathan al-diniyyah al-Islamiyyah. *Lughatu Ad-Dhat, 3*(2), 34–55. https://doi.org/10.37216/lughatuaddhat.v3i2.905

Zhou, X. (Emily), & Mann, S. (2021). Translanguaging in a Chinese university CLIL classroom: Teacher strategies and student attitudes. *Studies in Second Language Learning and Teaching, 11*(2), 265–289. https://doi.org/10.14746/ssllt.2021.11.2.5