



Arabic Learning in Early Childhood and Primary Education: A Scoping Review of Pedagogies, Literacy and Evidence Gaps

¹Abdul Basid, ²Nur Faizin, ³Habib Muharrom Sudarmawan, ⁴Moh. Fauzan, ⁵Abdul Adzim, ⁶Ahmad Munjin Nasih, ⁷Ali Ma'sum, ⁸Lilik Nur Kholidah, ⁹Achmad Sultoni, ¹⁰Mochammad Rizal Ramadhan, ¹¹Nur Mutamakkin, ¹²Mohammad Ahsanuddin, ¹³Nur Anisah Ridwan, ¹⁴Faris Khoirul Anam, ¹⁵Fery Fauzi, ¹⁶Muhammad Lukman Arifianto, ¹⁷Titit Thoriquityas

Universitas Negeri Malang, Indonesia

Corresponding Author: abdul.basid.fs@um.ac.id

Abstract

Arabic learning in early childhood involves different learner populations and outcomes, including spoken-to-standard literacy among native speakers and vocabulary learning among additional- and heritage-language learners. This study maps recent empirical evidence on Arabic learning for children approximately aged 3–12. Drawing on a JBI and PRISMA-ScR protocol framework, an initial publisher- and DOI-verified inventory of 23 primary-study candidates (2021–2026) was charted by population, design, pedagogy and outcome. Native-language studies examined diglossia, reading development and story-based supports; other studies concentrated on digital games, instructional media, teacher feedback and heritage access. Outcome measures ranged from platform feasibility to child-language performance, limiting direct comparison. The inventory is not a complete systematic-search sample: database exports, deduplication and independent screening remain outstanding. The preliminary evidence supports context-specific synthesis and calls for comparable, longitudinal and inclusive language measures.

Keywords: Arabic language; primary education; diglossia; scoping review

مستخلص البحث

يتناول تعلم اللغة العربية في الطفولة سياقات تعليمية ولغوية متباينة، منها اكتساب القراءة والكتابة لدى الأطفال الناطقين بالعربية، وتعلم المفردات لدى متعلمي العربية لغة إضافية، والمحافظة على العربية بوصفها لغة تراثية. يهدف هذا البحث إلى رسم خريطة أولية للدراسات التجريبية المتعلقة بتعليم العربية للأطفال بين الثالثة والثانية عشرة تقريباً. استخدم إطار منهجي مستند إلى إرشادات معهد جونا بريغز ومعايير الإبلاغ عن مراجعات النطاق، وجرى توصيف ثلاث وعشرين دراسة أولية مرشحة منشورة خلال المدة من 2021 إلى 2026. أبرزت الدراسات المتعلقة باللغة الأولى الازدواجية اللغوية وتنمية القراءة والسرديات القصصية، بينما تناولت دراسات أخرى الألعاب الرقمية والوسائط التعليمية والتغذية الراجعة والتعليم التراثي. اختلفت أدوات القياس بين صلاحية الوسائط والتحصيل اللغوي، مما يحذر من المقارنة المباشرة بين النتائج. ولا تمثل المجموعة الحالية حصيلة بحث منهجي شامل؛ فما زالت عمليات البحث في قواعد البيانات وفرز الدراسات بصورة مستقلة مطلوبة. وتبرز الحاجة إلى بحوث طويلة ومقارنة تراعي اختلاف الخلفيات اللغوية واحتياجات الأطفال.

الكلمات المفتاحية: اللغة العربية؛ التعليم الابتدائي؛ الازدواجية اللغوية؛ مراجعة النطاق

1. Introduction

1.1. General background and significance

Arabic education in childhood is not a single pedagogical problem. For many children, everyday interaction occurs in a spoken Arabic variety, whereas formal schooling introduces conventional written Standard Arabic. Other children approach Arabic as an additional foreign language in Islamic primary schools, as a heritage language in multilingual societies, or as a language of continuity amid forced migration. Teaching goals consequently span spoken vocabulary, oral comprehension, script recognition, phonological and morphological awareness, fluent reading, communicative participation, and language identity. Reviewing these outcomes together without identifying learners' prior language experiences risks producing a superficially unified but methodologically incoherent synthesis.

1.2. Relevant empirical research

Recent empirical work highlights the importance of the diglossic interface for first-language literacy. In a longitudinal kindergarten cohort, Standard Arabic linguistic abilities added predictive information to first-grade reading beyond spoken-language and orthographic skills (Asadi & Kavar, 2024). Morphological knowledge and listening comprehension have also been studied separately in spoken and literary varieties, underscoring why identical measurement labels need not indicate the same underlying construct across contexts (Asadi et al., 2024; Asadi & Kasperski, 2024). Studies of lexical distance and letter processing provide further reasons to resist transplanting general literacy techniques without attending to characteristics of Arabic (Asli-Badarneh & Asadi, 2023; Ghandour et al., 2025).

Meanwhile, educational technology features prominently in Arabic-as-additional-language settings. Examples include mobile vocabulary games for Grade 3 (Farida et al., 2022), Kahoot quizzes for Grade 5 (Zaki, 2025), and a game-based reading study with refugee children aged 5–10 (Tahir & Wang, 2024). These interventions differ fundamentally in what is measured: development feasibility, motivation, reading performance, or teacher feedback. A published systematic review of gamification among non-native Arabic learners provides broader context but was not designed to resolve childhood-specific differences across native, foreign, heritage, and displaced populations (Almelhes, 2024).

1.3. Evidence gaps and the present contribution

The rationale for a scoping rather than effectiveness review is therefore conceptual and evidential heterogeneity. Scoping reviews are suited to clarifying concepts, mapping types and distributions of evidence, and locating research gaps rather than estimating a single pooled effect (Munn et al., 2018; Pollock et al., 2024). This article develops a transparently qualified first-stage evidence map and a reproducible protocol for later comprehensive screening. It does not present publisher discovery as equivalent to a completed database-level systematic search.

1.4. Research questions and objectives

The review asks: (RQ1) Which learner groups, ages, countries and educational settings are represented? (RQ2) Which pedagogical approaches and Arabic-specific language features are addressed? (RQ3) Which academic, linguistic and

implementation outcomes are reported and through which designs? (RQ4) What limitations and underexamined research questions emerge, particularly when native-language literacy is compared with additional/heritage-language instruction?

1.5. Organization of the article

Section 2 sets out the reproducible eligibility and search protocol. Section 3 charts candidate evidence by language background and study design, interprets methodological contrasts, and develops a research agenda. Sections 4 and 5 explain limitations and state cautious conclusions.

2. Methods

2.1. Review design and reporting principles

The intended design follows the JBI guidance for scoping reviews and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) (Pollock et al., 2024; Tricco et al., 2018). The present version includes a planned protocol, a candidate study chart and an auditable claim–source appendix. It is not registered and must not be reported as a completed PRISMA-ScR study. No risk-of-bias threshold was used to exclude a candidate solely because it was observational or developmental; instead, design constraints are charted so that feasible media products are not described as causally effective interventions.

2.2. Population–Concept–Context (PCC) eligibility

Population: children broadly aged 3–12, including preschool and primary-grade cohorts; records in which the sampled age range marginally exceeds 12 would require full-text assessment to establish the proportion of eligible participants. Concept: empirical investigation of Arabic acquisition, emergent literacy, Arabic pedagogy, vocabulary or reading instruction, school-based Arabic learning materials, or Arabic heritage-language provision. Context: kindergartens, primary schools, family/community learning, migrant and refugee education, and heritage-language programs. Language status (L1 spoken/standard, additional/foreign, heritage) is a required extraction variable rather than an interchangeable synonym.

Candidate inclusion criteria were primary peer-reviewed journal research; publication between 1 January 2021 and 11 October 2026; a clearly relevant preschool/primary cohort or teachers serving one; Arabic-specific learning or language outcomes; and sufficient bibliographic metadata to locate a publisher or DOI record. Candidate exclusions were non-primary reviews, higher-education-only samples, studies of unrelated English learning, solely theological instruction without Arabic learning measures, editorials, duplicate reports, and sources whose relevance could not be substantiated from abstracts. Review papers are cited as methodological or contextual references, not charted as primary evidence.

2.3. Reproducible search strings and sources

The planned full search covers Scopus, Web of Science Core Collection, ERIC and relevant publisher/journal portals. A Scopus-ready core string is:

TITLE-ABS-KEY (("Arabic language" OR "learning Arabic" OR "teaching Arabic" OR "Arabic literacy" OR "Arabic reading" OR "Modern Standard Arabic" OR diglossia) AND (child* OR preschool* OR kindergarten* OR "primary school*" OR elementary OR "early childhood" OR "heritage language") AND (teach* OR learn* OR educat* OR pedagog* OR vocabular* OR read* OR literac* OR storytelling OR "digital game*" OR "language acquisition")) AND PUBYEAR > 2020 AND PUBYEAR < 2027

For Web of Science, the same conceptual groups should be implemented using TS= fields and document-type restrictions only after preserving an unfiltered export. The ERIC analogue should combine Arabic with descriptors such as early childhood education, elementary education, second language instruction, bilingual education and reading instruction. Backward and forward citation chasing should capture Arabic-language and Indonesian-language research that may not appear under English synonyms. Searches should be logged verbatim by platform, exact timestamp, coverage and filter sequence.

The work completed for this draft was a targeted web discovery and verification of publisher landing pages, DOI records and indexed abstracts on 11 October 2026. It identified 23 candidate original research reports that could be charted with sufficient source-specific detail. It did not yield a reproducible complete result export for Scopus, Web of Science or ERIC. Therefore database-specific record counts, duplication counts, reasons for full-text exclusion and final eligible-study numbers are explicitly left unreported. These 23 records are not asserted to be the final set of included studies.

2.4. Screening, data charting and synthesis plan

For the definitive review, two reviewers should independently screen title/abstract records, retrieve uncertain full texts, resolve disagreements through consensus or a third reviewer, and record exclusion reasons at full-text stage. DOI-first deduplication should be supplemented by normalized title, author and year matching, with multiple reports of the same cohort linked but not mechanically treated as independent participants. A reproducible charting sheet should capture citation and DOI; country; age/grade; L1/L2/heritage classification; sampled children/teachers; design and comparator; intervention duration and fidelity; linguistic variety; platform or teaching approach; outcome measure and timing; main reported numerical results; limitations; and whether the claim is based on abstract or inspected full text.

The planned synthesis uses descriptive counts only after screening completion; categorical mapping of populations and pedagogy; narrative comparison by language context; and an explicit separation of language-performance measures from motivation, usability and teacher-practice outcomes. Neither meta-analysis nor pooled effectiveness judgments are appropriate for the current heterogeneous and partially retrieved candidate set.

The PRISMA-ScR checklist cannot be marked fully complete for this manuscript: no defensible final study denominator, exclusion-reason distribution or database search log exists. Any submission must identify the article as a preliminary evidence map unless and until these records are produced.

Figure 1. PRISMA-ScR screening workflow. Missing database and screening counts are deliberately identified as pending.

Figure 1. PRISMA-ScR screening workflow - transparent draft status

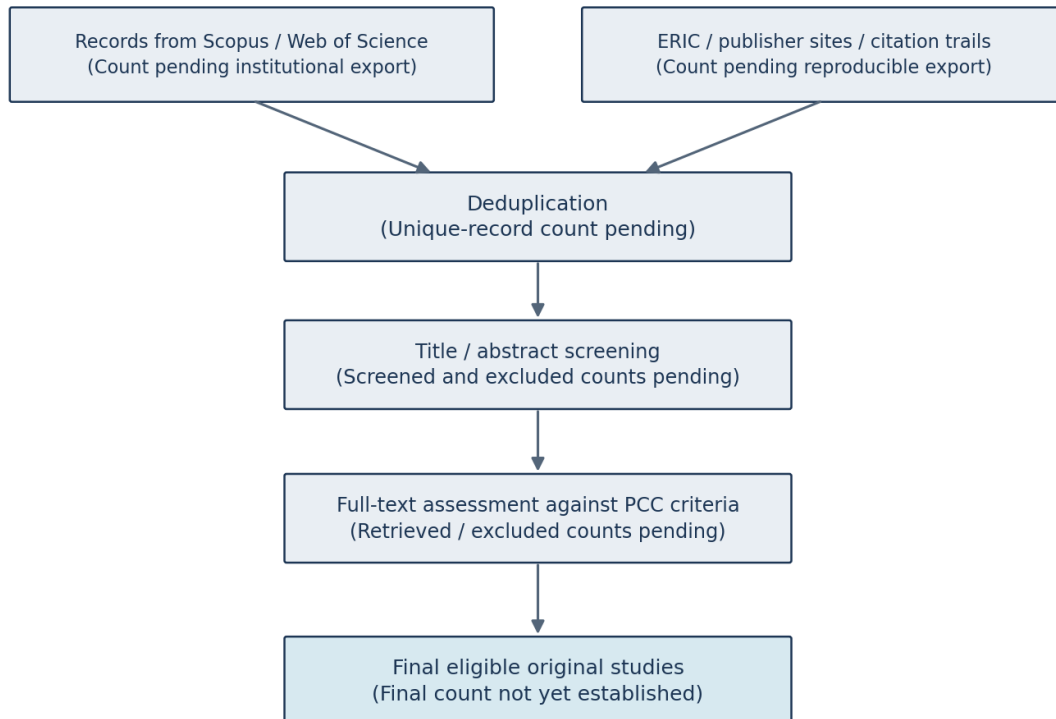


Figure 1 reproduces the identification, screening, eligibility and inclusion architecture of the supplied PRISMA template without borrowing its example counts. The candidate set of 23 is reported only as a verified source inventory for drafting. It cannot be used as a final PRISMA inclusion box before database export, deduplication and independent screening.

Figure 1 makes clear that the 23 report candidates cannot be used to populate any final inclusion box without an export log and independent screening.

3. Results and Discussion

3.1. Native-language literacy and the spoken–standard interface

Table 1 charts 13 initial primary studies of first-language Arabic literacy and development. The studies concentrate on spoken–standard linguistic distance, kindergarten predictors of reading, story-based intervention, morphological and orthographic learning, and teacher identification of reading difficulties. Multiple papers use a 261-child kindergarten cohort (Asadi & Kavar, 2024; Asadi et al., 2024; Asadi & Kasperski, 2024); until underlying participant overlap is audited, treating their sample sizes as additive would falsely inflate apparent coverage.

Table 1. Publisher/DOI-verified original-study candidates: native Arabic literacy

Study	Participants / setting	Design or instructional focus	Evidence charted and interpretive caution
Asadi, Atila, & Saleh (2024)	102 Arabic-speaking kindergarten children; first-grade follow-up	Three-month interactive story-reading intervention; textbook-derived vs familiar stories	Intervention group outperformed comparison on practiced and unfamiliar reading items. Caution: Two-group setting; long-term transfer beyond Grade 1 untested.
Asadi & Kavar (2024)	261 kindergartners followed to Grade 1	Longitudinal predictors: spoken/standard language; phonemic, syntactic and orthographic skills	Standard Arabic language index predicted first-grade reading after controls. Caution: Predictive association, not an instructional-effect estimate.
Asadi & Asli-Badarneh (2023)	137 children, first to second grade	Longitudinal word-reading assessment by spoken–standard lexical distance	Reading acquisition varied by lexical relations between spoken and standard Arabic. Caution: Observational; no teaching intervention.

Study	Participants / setting	Design or instructional focus	Evidence charted and interpretive caution
Asli-Badarneh & Asadi (2023)	180 native Arabic-speaking Grade 1–2 children	Word lists and texts varying lexical/phonological proximity	Reading accuracy/fluency were examined by lexical-distance condition. Caution: Mechanistic association; not classroom treatment.
Asadi et al. (2024; morphology)	261 kindergarten children followed to Grade 1	Inflectional morphology in spoken and standard varieties linked to literacy	Study charted morphology in two varieties and subsequent reading/spelling. Caution: May overlap with other published 261-child cohorts; independence must be checked.
Asadi & Kasperski (2024)	261 kindergartners followed to Grade 1	Longitudinal simple-view-of-reading models by language variety	Reported models explained 52% (spoken) and 48% (literary) of variance. Caution: Explained variance is not a treatment effect.
Mansour-Adwan et al. (2023)	Arabic-speaking kindergartners followed to Grade 1	Longitudinal linguistic profiles and subsequent reading	Lower early linguistic skills were associated with later reading difficulty. Caution: Predictive study rather than intervention evidence.
Shahbari-Kassem et al. (2024)	139 kindergarten children (71 intervention; 68 comparison)	12-week scaffolded stories progressing from spoken to standard forms	Intervention related to better story comprehension and vocabulary; all-spoken stories scored highest by condition. Caution: Findings depend on the particular texts and diglossic distance sequence.

Study	Participants / setting	Design or instructional focus	Evidence charted and interpretive caution
Haj et al. (2026)	403 kindergartners (250 intervention; 153 control)	15-week, thrice-weekly diglossia-centered multimodal intervention	Reported larger gains in language, emergent literacy, EF and letter knowledge. Caution: A publisher correction appeared 17 September 2026; use corrected record.
Balshe & Masterson (2026)	116 monolingual third-grade children	Paired-associate word learning with vs without orthographic cues	Written forms aided novel-word learning; dense consonant diacritics posed difficulty. Caution: Laboratory word-learning task; classroom generalizability unknown.
Hassanein et al. (2023)	School-age Arabic readers	Modeling phonological, morphological, and orthographic processes	Study compared reading-skill predictors in Arabic. Caution: No intervention; classroom implications indirect.
Khatib-Abbas & Lipka (2025)	Arabic-speaking Grade 3 learners and their teachers	Pilot study of teacher accuracy identifying fluency/comprehension problems	Examined identification practices and referral accuracy. Caution: Pilot, setting-specific and not a treatment.
Ghandour et al. (2025)	Arabic-speaking beginning readers	Visual-attention span, letter knowledge, phonological-awareness predictors	Visual-attention association with reading was mediated by letter knowledge. Caution: Correlational findings should not be called intervention results.

The chart in Table 1 shows that one type of primary evidence concerns predictors, for example the association of kindergarten spoken/standard language abilities with first-grade reading (Asadi & Kawar, 2024), whereas another concerns a specified educational treatment. In a 102-child intervention, kindergarten story-reading activities were followed by differences on practiced and unfamiliar first-grade reading items (Asadi, Atila, & Saleh, 2024). A separate 139-child storybook intervention manipulated the

distance between spoken and standard forms, with vocabulary and comprehension outcomes reported (Shahbari-Kassem et al., 2024). These results provide mechanistic and pedagogical starting points rather than a pooled story-reading effect.

One recent multi-domain ADMIN trial enrolled 403 kindergartners, with 250 assigned to intervention and 153 to control, and reported larger gains in language, early literacy and executive-function-related measures over 15 weeks (Haj et al., 2026). The publisher issued a correction on 17 September 2026 (<https://doi.org/10.1007/s11145-026-10905-2>), so its final outcome tables and detailed estimates require verification against the corrected version before submission. Orthographic experiments with 116 third-graders suggest that presenting print alongside novel spoken word forms can support learning, but also reveal complexity from consonant diacritics (Balshe & Masterson, 2026).

The publisher's correction to Haj et al. (2026) additionally changes the upper bound of a reported control-group effect-size range from $d = 0.81$ to $d = 0.65$, and another effect-size range from $d = 0.86$ to $d = 0.75$. These values must be read from the corrected record rather than superseded descriptions. Its Table 4 also adds significance marks to 6.41 (Haj et al., 2026, correction).

3.2. Primary pedagogy, refugee education and heritage contexts

Table 2 charts 10 initial primary studies focused on Arabic instructional materials, refugee or heritage contexts, and teacher feedback. Native/additional language status is variably specified among these records. Among these records, digital and game-based approaches recur, but the outcomes do not automatically index the same construct. A viable software product, a favorable motivation response and an increased vocabulary score represent different levels of evidence.

Table 2. Publisher/DOI-verified original-study candidates: primary-school pedagogy and heritage contexts (10 records; mixed language status; not a final PRISMA sample).

Study	Participants / setting	Design or instructional focus	Evidence charted and interpretive caution
Tahir & Wang (2024)	30 refugee children aged 5–10 in informal learning	One-group pretest–posttest Arabic reading game; usability data	Reading scores increased; usability varied with learner characteristics. Caution: No comparison group; measured gain not attributable uniquely to game.

Study	Participants / setting	Design or instructional focus	Evidence charted and interpretive caution
Bohnacker et al. (2025)	Heritage-language education serving preschool/primary-age learners in Sweden	Policy/literature/statistical account combined with two empirical studies	Charts institutional opportunities and constraints in Arabic heritage education. Caution: Context-focused; not a measured child-language intervention.
Alghamdi (2025)	40 male Arabic teachers teaching Grade 5, Saudi Arabia	Mixed-method teacher training and interactional feedback during digital games	Greater change in teachers' feedback practices, not direct student-proficiency proof. Caution: Teacher behavior is primary outcome; limited learner language results.
Azizah & Batubara (2021)	Grade 1 Islamic elementary school learners, Indonesia	Development and feasibility assessment of Arabic-vocabulary quiz application	Feasibility and user-acceptability evidence reported. Caution: Feasibility does not establish sustained language gains.
Farida et al. (2022)	Grade 3 primary learners, Indonesia	Mobile Arabic-vocabulary game; development validation	Expert feasibility: 98.75% media and 81.25% content; user suitability 84.65%. Caution: Feasibility indices should not be equated with effectiveness.
Ritonga et al. (2022)	30 Grades 1 and 3 pupils; 4 teachers, Indonesia	Qualitative questionnaires and teacher interviews about language games	Participants described word/animal guessing games as enjoyable. Caution: Perceived engagement; does not causally measure vocabulary improvement.

Study	Participants / setting	Design or instructional focus	Evidence charted and interpretive caution
Rahmah & Fauji (2024)	20 fourth-grade learners, Indonesia	Pre-post classroom use of Powtoon web media	Reported mean N-gain 0.515 (moderate) for vocabulary. Caution: Very small classroom sample; unclear counterfactual.
Zaki (2025)	Fifth-grade Ibtidaiyah Indonesia	Madrasah learners, Pretest-posttest vocabulary evaluation Kahoot	Mean vocabulary score rose from 30.90 to 80.00. Caution: Unclear equivalent comparison group; effect attribution limited.
Faizah et al. (2024)	Grade 4 Islamic primary pupils, Indonesia	Educational-game development with Match activities Make a	Developed and evaluated learning-media product for vocabulary. Caution: Development design; causal inference requires further trials.
'Alawiyyah et al. (2026)	Grades 1–6 elementary learners, Indonesia	ADDIE design of Arabic digital-comic media with QR links	Arabic-specific media development/evaluation in elementary school. Caution: Development and test settings; outcomes require independent replication.

Several studies in Table 2 report engagement or technology feasibility rather than language effects. A Grade 3 mobile-game development study reported media feasibility of 98.75%, subject-content feasibility of 81.25% and learner product suitability of 84.65% (Farida et al., 2022). Those ratings evaluate design acceptability; they do not measure durable Arabic acquisition. A qualitative study of 30 pupils and four teachers similarly points to enthusiasm for guessing games rather than establishing a causal gain in vocabulary (Ritonga et al., 2022). By contrast, a Grade 5 Kahoot evaluation reported a vocabulary mean increase from 30.90 to 80.00; without a clearly equivalent control group, the reported pre-post difference should be interpreted as within-setting change, not a general effect estimate (Zaki, 2025).

Informal and heritage settings broaden the mapping beyond conventional classrooms. Tahir and Wang (2024) evaluated Arabic reading through a game with 30 refugee children aged 5–10 and reported higher post-use scores, while noting age and experience

effects on usability. Bohnacker et al. (2025) examined Swedish Arabic heritage education through provision statistics and empirical studies, drawing attention to institutional conditions rather than one standardized child outcome. Training-focused work by Alghamdi (2025), involving 40 primary teachers, demonstrates that the proximal outcome of educational technology may be teachers' feedback repertoire, not directly measured child proficiency. These different outcome positions are important for the argument developed in Section 4.

Integrated interpretation of candidate evidence

3.3. Distinct learner populations and evidence streams

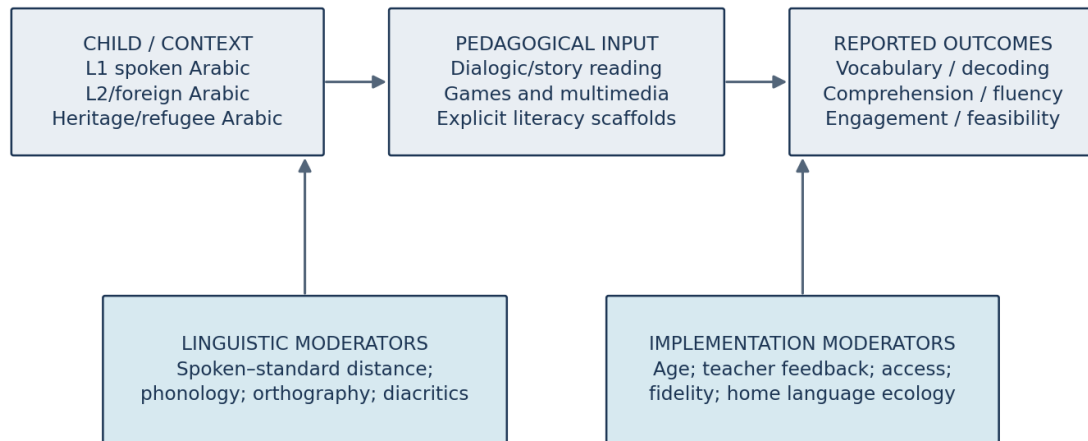
The most defensible provisional conclusion is that the phrase 'Arabic for children' captures multiple evidence domains requiring explicit analytical stratification, including studies that do not clearly specify a learner's language status. For native-speaking children, the educational challenge often involves moving between spoken vernacular and school-based Standard Arabic literacy. Studies of lexical/phonological distance and longitudinal reading predictors address this transition directly (Asadi & Asli-Badarneh, 2023; Asli-Badarneh & Asadi, 2023). For non-native children in primary school, the dominant instructional question in the identified studies often concerns learning new vocabulary and engaging with unfamiliar Arabic forms through games or multimedia (Azizah & Batubara, 2021; Farida et al., 2022; Rahmah & Fauji, 2024). Heritage and refugee cases add distinctive institutional, identity and continuity constraints (Bohnacker et al., 2025; Tahir & Wang, 2024).

Figure 2 locates those groups, pedagogical supports and outcomes within a shared but conditional framework. The model deliberately distinguishes linguistic moderating factors (diglossic distance and orthographic complexity) from implementation conditions (teacher feedback, access and home-language ecology). It is a conceptual integration of the mapped studies, not a causal pathway estimated from pooled data.

As Figure 2 further emphasizes, a change in home-language ecology or access to qualified teachers can alter what the same instructional activity means across native, foreign and heritage settings.

Figure 2. Context-sensitive conceptual organization of childhood Arabic-learning evidence.

Figure 2. A context-sensitive framework for Arabic learning in childhood



Conceptual organization of candidate evidence, not an estimated causal model

3.4. Spoken–standard bridging and instructional design

Arabic diglossia should not be reduced to the assumption that children's spoken language constitutes a deficiency. Rather, studies can operationalize the relationship between known spoken items and less familiar standard forms, then ask whether carefully graded exposure aids representation of literary vocabulary, reading and comprehension. The graded-story intervention reported by Shahbari-Kassem et al. (2024) and the later ADMIN work (Haj et al., 2026) exemplify such program design. In parallel, Asadi and Kasperski (2024) show that the relative contributions of decoding and listening comprehension depend on which language variety is measured. The implication is not that one universal diglossia method is proven superior; it is that research instruments, stimuli and pedagogy must specify the language variety tested.

Visual script characteristics add a further constraint. Arabic word reading draws on orthographic, morphological and phonological processing, as addressed in predictor analyses by Hassanein et al. (2023). Ghandour et al. (2025) found that letter knowledge mediated the association between visual-attention span and early reading. In a different experimental format, Balshe and Masterson (2026) found an advantage of visible written forms for novel-word learning among third-graders while also observing difficulty related to consonant-diacritic density. These are plausible instructional mechanisms but are not substitute evidence for a controlled classroom curriculum.

3.5. Digital engagement versus language learning

A recurring methodological risk in digital-Arabic studies is outcome substitution: a product is judged engaging, so readers conclude that it improves acquisition. Table 2 prevents that inference by charting media validity, affective responses, teacher change and measured language performance separately. Development studies (Azizah & Batubara, 2021; Faizah et al., 2024) may justifiably conclude that a platform has passed specified design checks. They cannot establish sustained vocabulary retention without a comparison, validated achievement measure and follow-up. Likewise, qualitative descriptions of games may indicate acceptability without determining a treatment effect (Ritonga et al., 2022).

Even a statistically reported pre-post improvement requires attention to classroom assignment, baseline variation, alternative instruction, exposure time and assessment alignment. The Kahoot report (Zaki, 2025) and the 20-pupil Powtoon report (Rahmah & Fauji, 2024) supply potentially useful numerical outcomes; however, these local outcomes do not make competing digital tools directly comparable. The refugee game study likewise used a one-group pretest-posttest design, restricting causal interpretation while yielding valuable implementation insights for vulnerable learners (Tahir & Wang, 2024).

3.6. Pedagogical implications

Three design principles follow provisionally. First, research should specify whether children are building initial spoken Arabic, bridging from spoken to standard Arabic, strengthening literacy in a heritage environment, or recovering interrupted Arabic learning. Second, tasks should be aligned with age and target construct: oral language and story comprehension, script/word decoding, vocabulary production, or meaningful use of Arabic should not be conflated. Third, technology evaluations should assess linguistic progress separately from engagement, accessibility, and teacher practices. A well-designed game may provide repeated practice or prompt more interaction, but the pathway from interface to learning remains a hypothesis requiring direct measurement.

These principles also caution against treating every local school study as equally generalizable. Preschool interventions examining spoken-standard transfer (Asadi, Atila, & Saleh, 2024; Haj et al., 2026) answer a different question from a Canva digital-comic development study spanning Grades 1–6 (‘Alawiyyah et al., 2026). Additionally, evidence from a single country or school type should not be assumed to transfer to multilingual heritage settings where teachers, curriculum and instructional time differ (Bohnacker et al., 2025).

3.7. Research priorities

Figure 3 turns the limitations charted in Tables 1–2 into research priorities. These include transparent linguistic stratification, stronger comparison designs, delayed outcomes, inclusive recruitment and distinction between technological design and actual language attainment. A future completed systematic search may change the relative prominence of individual priorities. The figure should therefore be interpreted as a hypothesis-generating agenda rather than a quantitatively ranked gap map.

Figure 3 therefore summarizes a research-design agenda derived from the mapped limitations rather than rankings supported by prevalence estimates.

Figure 3. Priority design and reporting gaps inferred from initial study mapping.

Figure 3. Priority evidence gaps and recommended study designs

Cross-context comparability	Report L1/L2/heritage status and dialect; standardize learner age
Intervention quality	Use controlled comparison, preregistered outcomes and fidelity measures
Learning beyond novelty	Add delayed vocabulary, reading, and oral-production assessments
Equity / accessibility	Disaggregate by disability, home language and digital access
Mechanism clarity	Separate software feasibility, engagement and language attainment

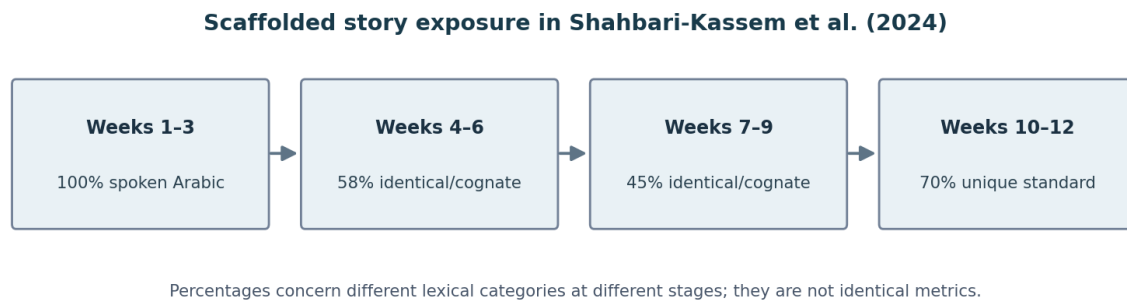
Research agenda inferred from heterogeneity and limitations of identified candidate studies

Future studies could test whether structured spoken-to-standard story sequences yield gains in untrained vocabulary and reading beyond immediate test items; whether games add value over time-matched active instruction; how teacher corrective feedback mediates oral use of Arabic; and how heritage, disability and refugee-status factors shape accessibility. Multi-site trials with validated parallel measures, intervention fidelity, and follow-up would help distinguish context-specific enthusiasm from durable skill development. For the scoping review itself, full retrieval and independent screening remain prerequisites before any claim about global evidence volume or geographic underrepresentation can be made.

3.8. Visual cross-study synthesis

Figure 4 depicts the 12-week story sequence described by Shahbari-Kassem et al. (2024). Its stages represent changing lexical compositions of the material, not a single dose-response axis: the percentages refer to different spoken, cognate or standard word classes. Figure 4 therefore illustrates pedagogical sequencing rather than evidence that steadily increasing Standard Arabic always produces higher comprehension.

Figure 4. Story-language progression in the 139-child intervention, reconstructed from the authors' described instructional sequence (Shahbari-Kassem et al., 2024).



As the sequence in Figure 4 clarifies, familiarity and linguistic distance should be separated analytically from general exposure duration. Figure 5 groups the ten pedagogy/heritage records in Table 2 by their principal evaluation focus. Four records mainly validated learning media; three reported child-language pre-post outcomes; the remaining records examined engagement, teacher feedback or heritage provision.

Figure 5. Descriptive classification of the ten Table 2 candidate studies by their main evidence object; not a count of eligible studies after PRISMA screening.

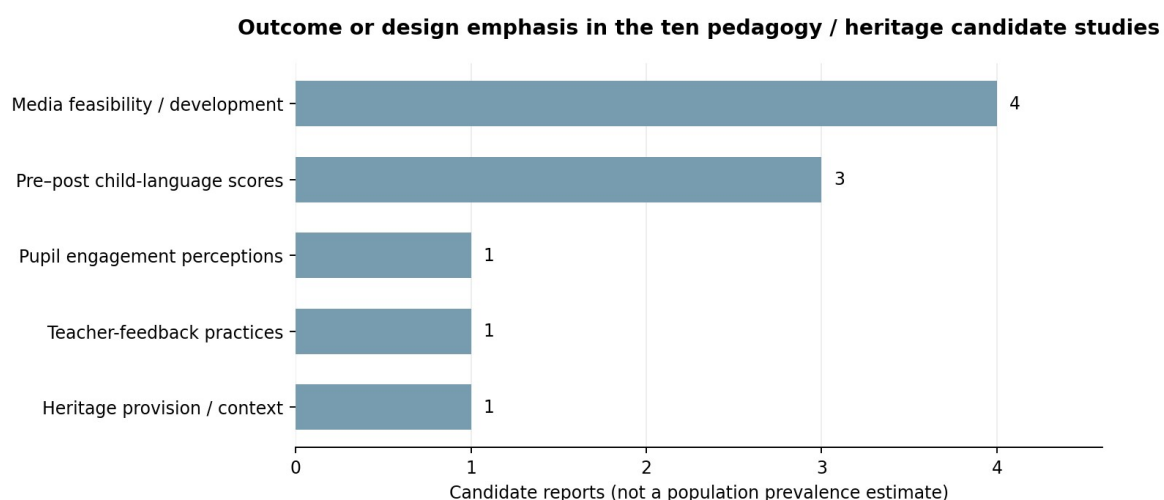
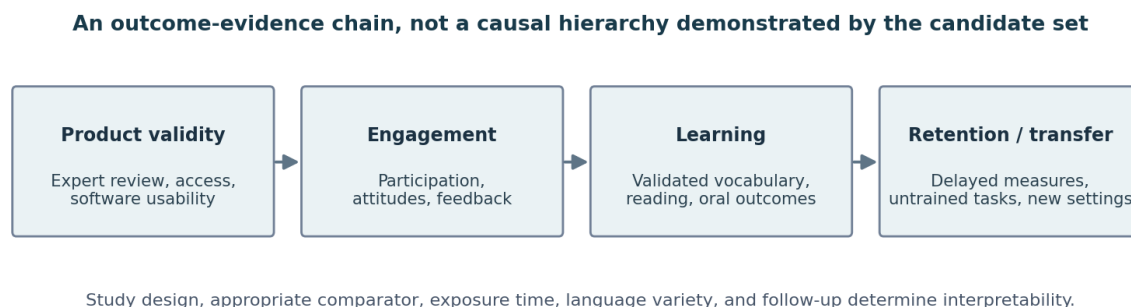


Figure 5 emphasizes why counting favorable studies is misleading when feasibility and learning are assigned the same status. Figure 6 consequently proposes a reporting chain that distinguishes product validity, engagement, immediate language learning and retention/transfer. The diagram is a methodological organizing device, not a demonstrated causal sequence.

Figure 6. Proposed evaluation chain for technology-enhanced Arabic pedagogy, integrating evidence types discussed in Table 2.



A future study guided by Figure 6 would measure language gains independently of product acceptance, use an active comparator where feasible, and reassess retention after instruction. Figure 6 also highlights an empirical gap identified across the candidate sources: digitally mediated learning is frequently described without robust evidence of delayed transfer to untrained vocabulary, reading or spoken-language tasks.

4. Limitations and strengths

This manuscript is deliberately an initial scoping draft, not a finished review. Searches were conducted through targeted online discovery and bibliographic verification rather than full reproducible bulk retrieval from each database. The 23 mapped primary records are an auditable starting inventory; they are not a comprehensive denominator. A true PRISMA-ScR flow requires source-specific counts, date-stamped export files, DOI/title deduplication, an inclusion/exclusion log, independent screening, full-text retrieval and reasons for exclusion. No fabricated counts have been substituted for these steps.

A second limitation arises from available source granularity. Many claim verifications rest on authoritative publisher abstract text rather than examination of every instrument, table or supplementary file. Accordingly, the discussion reports only abstract-supported numerical details and cautiously qualifies causal language. Several studies may share overlapping cohorts (especially reports involving 261 kindergarteners), and many additional-language studies were conducted in one local educational system. Geographic and publication-language representativeness therefore remain undetermined.

Strengths include explicit PCC boundaries, identification of major language-status strata, careful separation of language gains from feasibility and engagement, transparent acknowledgement of a 2026 published correction, and a claim–source ledger in Appendix A. These features support a coherent full review once its systematic search and screening records are available.

Appendix Table A1 supplies a claim-to-reference audit trail, but abstract-level checks cannot replace a complete re-examination of source methods and outcomes.

5. Conclusion

The initial evidence map indicates that Arabic teaching and learning among children is best understood as a context-sensitive intersection of language status, literacy demands and educational design. Native Arabic literacy studies emphasize the spoken–standard interface, early predictors of reading, structured stories and orthographic mechanisms. Primary-school teaching and additional/heritage-language studies contribute evidence about games, multimedia, teacher interaction and educational provision, but frequently employ designs that cannot isolate treatment effects. A publication-ready scoping review should not infer a single best method from these heterogeneous sources. Its central contribution should be a transparent map of who was studied, what was taught, which outcomes were actually measured and where appropriately controlled, inclusive and longitudinal studies are still needed.

References

- ‘Alawiyyah, A. K., Murtadho, N., & Nurhidayati, N. (2026). Visualizing language: A Canva-based digital comic for enhancing Arabic learning in elementary education. *Tsaqofiya: Jurnal Pendidikan Bahasa dan Sastra Arab*, online publication. <https://doi.org/10.21154/b2mpsz98>
- Alghamdi, A. (2025). Investigating the effectiveness and typology of teachers’ interactional feedback in supporting digital game-based language learning. *Frontiers in Education*, 10, Article 1645840. <https://doi.org/10.3389/feduc.2025.1645840>
- Almelhes, S. A. (2024). Gamification for teaching the Arabic language to non-native speakers: A systematic literature review. *Frontiers in Education*, 9, Article 1371955. <https://doi.org/10.3389/feduc.2024.1371955>
- Asadi, I. A., & Asli-Badarneh, A. (2023). The impact of diglossia-effect on reading acquisition among Arabic-speaking children: A longitudinal study. *Journal of Psycholinguistic Research*, 52(5), 1919–1937. <https://doi.org/10.1007/s10936-023-09969-w>
- Asadi, I. A., & Kasperski, R. (2024). Reading comprehension in the Arabic diglossia: The SVR between the spoken and literary varieties. *Journal of Psycholinguistic Research*, 53, Article 27. <https://doi.org/10.1007/s10936-024-10054-z>
- Asadi, I. A., & Kwar, K. (2024). Learning to read in Arabic diglossia: The relation of spoken and standard Arabic language in kindergarten to reading skills in first grade. *Literacy Research and Instruction*, 63(4), 360–376. <https://doi.org/10.1080/19388071.2023.2217274>
- Asadi, I. A., Asli-Badarneh, A., Ibrahim, R., & Hamzah, H. (2024). The impact of diglossia on inflectional morphological constructions and their relation to literacy skills: A longitudinal study. *Journal of Speech, Language, and Hearing Research*, 67(2), 496–510. https://doi.org/10.1044/2023_JSLHR-23-00318
- Asadi, I. A., Atila, N., & Saleh, S. (2024). The Arabic diglossia reality: The effect of specific story reading in kindergarten on reading acquisition in first grade. *Literacy Research and Instruction*, 63(4), 342–359. <https://doi.org/10.1080/19388071.2023.2208187>

- Asli-Badarneh, A., & Asadi, I. A. (2023). The impact of lexical and phonological distance on reading acquisition: The diglossic context of Arabic. *Journal of Research in Reading*, 46(4), 376–392. <https://doi.org/10.1111/1467-9817.12433>
- Azizah, F. N., & Batubara, H. H. (2021). Pengembangan aplikasi edukasi kuis Bahasa Arab sebagai alternatif media belajar siswa kelas I Madrasah Ibtidaiyah. *Alsina: Journal of Arabic Studies*, 3(2), 219–240. <https://doi.org/10.21580/alsina.3.2.8432>
- Balshe, R., & Masterson, J. (2026). Word learning in Arabic-speaking primary school students. *European Journal of Psychology of Education*, 41, Article 13. <https://doi.org/10.1007/s10212-025-01061-2>
- Bohnacker, U., Wehbe, P., Daravigka, K., & Haddad, R. (2025). Arabic heritage language education in Sweden: Opportunities and challenges. *European Educational Research Journal*, advance online publication. <https://doi.org/10.1177/14749041251318150>
- Faizah, F., A'yuni, K., & Al-Hakim, M. F. (2024). Development of the Alibaba media game assisted with the Make a Match model in Arabic language learning to improve the vocabulary mastery of class IV primary school students. *Al-Mudarris: Journal of Education*, 7(1), 55–65. <https://doi.org/10.32478/kpve1947>
- Farida, H., Ulfa, S., & Kuswandi, D. (2022). Pengembangan mobile game based learning kosakata Bahasa Arab untuk siswa kelas 3 sekolah dasar. *Edcomtech: Jurnal Kajian Teknologi Pendidikan*, 7(1), 38–50. <https://doi.org/10.17977/um039v7i12022p38>
- Ghandour, A., Trouche, E., Guillo, D., & Valdois, S. (2025). The influence of visual attention on letter recognition and reading acquisition in Arabic. *Frontiers in Psychology*, 16, Article 1628051. <https://doi.org/10.3389/fpsyg.2025.1628051>
- Haj, L., Schiff, R., Ghawi-Dakwar, O., & Saiegh-Haddad, E. (2026). Investigating the role of Arabic diglossia-centered multi-domain intervention (ADMIN) in enhancing Palestinian-Arabic speaking kindergarten children's emergent literacy. *Reading and Writing*, advance online publication. <https://doi.org/10.1007/s11145-026-10766-9>
- Haj, L., Schiff, R., Ghawi-Dakwar, O., & Saiegh-Haddad, E. (2026). Correction: Investigating the role of Arabic diglossia-centered multi-domain intervention (ADMIN) in enhancing Palestinian-Arabic speaking kindergarten children's emergent literacy. *Reading and Writing*. <https://doi.org/10.1007/s11145-026-10905-2>
- Hassanein, E. E. A., Johnson, E. S., Ibrahim, S., & Alshaboul, Y. (2023). What predicts word reading in Arabic?. *Frontiers in Psychology*, 14, Article 1077643. <https://doi.org/10.3389/fpsyg.2023.1077643>
- Khatib-Abbas, S., & Lipka, O. (2025). Teacher identification of reading difficulties among Arabic-speaking third graders in Israel: A pilot study. *Annals of Dyslexia*, 75, 262–288. <https://doi.org/10.1007/s11881-025-00331-4>
- Mansour-Adwan, J., Shalhoub-Awwad, Y., Cohen-Mimran, R., & Khateb, A. (2023). Reading in kindergarten Arabic-speaking children with low linguistic

- skills: A longitudinal study. *Applied Psycholinguistics*, 44(2), 265–291. <https://doi.org/10.1017/S014271642300019X>
- Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18, Article 143. <https://doi.org/10.1186/s12874-018-0611-x>
- Pollock, D., Peters, M. D. J., Tricco, A. C., Munn, Z., Jia, R. M., Alexander, L., Pieper, D., Evans, C., Godfrey, C. M., Brandão de Moraes, É., Saran, A., Campbell, F., & Khalil, H. (2024). Scoping reviews. In E. Aromataris, C. Lockwood, K. Porritt, B. Pilla, & Z. Jordan (Eds.), *JBİ manual for evidence synthesis*. JBI. <https://doi.org/10.46658/JBİMES-24-09>
- Rahmah, H., & Fauji, I. (2024). Enhancing Arabic vocabulary mastery in primary students through Powtoon-based web media. *Indonesian Journal of Islamic Studies*, 12(2). <https://doi.org/10.21070/ijis.v12i2.1692>
- Ritonga, M., Ritonga, A. W., Wahyuni, S., & Julhadi, J. (2022). Language game as an alternative model to improve Arabic vocabulary ability. *Ijaz Arabi Journal of Arabic Learning*, 5(3). <https://doi.org/10.18860/ijazarabi.v5i3.17123>
- Shahbari-Kassem, A., Asli-Badarneh, A., Hende, N., & Roby-Bayaa, A. (2024). Reading stories in Arabic: The impact of lexico-phonological and diglossic distance level on comprehension and receptive and productive vocabulary among Arab kindergarten children. *Frontiers in Education*, 9, Article 1394024. <https://doi.org/10.3389/feduc.2024.1394024>
- Tahir, R., & Wang, A. I. (2024). Evaluating the effectiveness of game-based learning for teaching refugee children Arabic using the integrated LEAGUE-GQM approach. *Behaviour & Information Technology*, 43(1), 110–138. <https://doi.org/10.1080/0144929X.2022.2156386>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- Zaki, M. (2025). Enhancing Arabic vocabulary mastery through Kahoot: Evidence from elementary madrasah students. *Al-Ta'rib: Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab*, 13(2), 423–442. <https://doi.org/10.23971/altarib.v13i2.10544>