



Arabic Essay Writing among Non-Native Learners: A Scoping Review of Pedagogies, Feedback, Assessment, 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, ¹⁷Titis Thoriquattyas

Universitas Negeri Malang, Indonesia

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

Abstract

Arabic essay writing requires organization, linguistic accuracy, feedback, and revision, yet the literature uses heterogeneous tasks and outcomes. This article develops a scoping-review framework for mapping essay and paragraph writing among non-native Arabic learners. Population–Concept–Context criteria, multilingual Boolean searches, and PRISMA-ScR reporting fields were specified. A targeted preliminary register of 20 original-study candidates was charted, without comprehensive Scopus, Web of Science, or ERIC screening. Candidate studies examine process writing, genre scaffolding, peer feedback, collaborative tasks, digital resources, error analysis, and assessment. Their outcomes range from independently assessed texts to orthographic errors and learner perceptions. These outcomes cannot support a common effectiveness estimate. A completed review will require independent eligibility decisions, full-text extraction, comparable writing measures, and reconciled PRISMA counts. The present synthesis is preliminary and does not establish field-wide research frequencies.

Keywords: Arabic composition; Essay writing; Writing assessment; Scoping review

مستخلص البحث

تتناول هذه الورقة تعليم كتابة المقالة باللغة العربية لدى المتعلمين غير الناطقين بها، مع التركيز على التخطيط وتنظيم الأفكار والصحة اللغوية والتغذية الراجعة والمراجعة. وتهدف إلى بناء إطار لمراجعة استطلاعية ترصد طرائق التدريس وأساليب التقويم وحدود الأدلة المتاحة. حُدثت معايير السكان والمفهوم والسياق، وصيغت استراتيجيات بحث متعددة اللغات وفق إرشادات المراجعات الاستطلاعية. وجرى إعداد سجل أولي موجه يضم عشرين دراسة بحثية مرشحة، من دون استكمال البحث الشامل في قواعد البيانات أو إجراءات الفرز المستقلة. وتشمل الدراسات المرشحة الكتابة العملية، وتعليم الأنواع النصية، والتعاون، والتصحيح، والموارد الرقمية، وتحليل الأخطاء. وتختلف المخرجات بين جودة النص المكتوب ودقة الإملاء وآراء المتعلمين؛ لذلك لا

يجوز عدها مقاييس متكافئة لفاعلية التدريس. وتتطلب المراجعة النهائية التحقق من النصوص الكاملة وتوثيق قرارات الإدراج وأعداد مخطط برزما. وعليه تظل النتائج الحالية أولية ومحدودة النطاق. الكلمات المفتاحية: التعبير الكتابي؛ كتابة المقالة؛ تقييم الكتابة؛ مراجعة استطلاعية

1. Introduction

1.1. Background and significance

Writing an Arabic essay is not equivalent to reproducing words or sentences. A complete composition requires topic interpretation, the generation and ordering of ideas, appropriate lexical and morphological choices, cohesive linking, register awareness and revision. These skills interact but are not interchangeable. In additional-language contexts, proficiency in Arabic script, grammatical inflection and target variety may shape the time available for planning and higher-order textual decisions. Research therefore needs to distinguish sentence- or word-level accuracy from the ability to sustain a coherent essay.

1.2. Previous empirical research

The available primary literature suggests multiple educational routes into composition. A university essay-writing intervention explicitly foregrounded prewriting and successive writing stages (Almugim, 2023). An activity-based study instead examined Arabic paragraph writing among 25 undergraduate Arabic-as-a-foreign-language learners (Sathikulameen et al., 2023). At another level, a survey of 25 secondary pupils recorded favourable perceptions of collaborative strategic reading as preparation for Arabic composition (Azman and Bin Chik, 2024). These studies address related tasks but represent different evidentiary objects: assessed texts, pre/post questionnaire responses and learners' reported beliefs.

Recent research also connects composition with feedback and linguistic form. In a nonequivalent classroom comparison of Think–Pair–Share, posttest means were reported as 77.66 versus 58.81, with an experimental-group normalized gain of 0.55 (Amaliyah et al., 2025). Peer-editing classroom action research recorded an increase in mastery from 12 to 32 of 36 learners over successive cycles, but concentrated on numeral–noun agreement rather than complete essays (Sa'diyah et al., 2026). The existence of both broader composition outcomes and narrower form-focused outcomes raises a critical interpretive question: what exactly has improved when authors refer to 'Arabic writing skills'?

1.3. Gap and distinct contribution

Available evidence also highlights measurement and access. A descriptive study of non-native university students analyzed grammatical and orthographic errors in essay tests (Al-Adwan and El-Shara, 2025). Texts and self-reports used within a complexity–accuracy–fluency framework identified difficulties in syntax, accuracy and sustained idea development (Syaifullah et al., 2026). In contrast, online-discussion research described mixed learner responses and uneven writing development rather than a clear intervention effect (Anwar et al., 2025). A rigorous mapping exercise must chart both positive and inconclusive reports and must not translate learner enthusiasm automatically into demonstrated linguistic proficiency.

1.4. Research objective and contribution

The review is intended to contribute a writing-specific evidence taxonomy that separates (a) the unit of production, from orthographic response to paragraph and essay; (b) instructional process, including prewriting, drafting and feedback; (c) measurement, including perception, textual quality and linguistic error coding; and (d) the inference permitted by each design. A scoping approach fits this heterogeneity and the aim of locating evidence gaps rather than estimating a single pooled effect (Munn et al., 2018). The completed report would follow the PRISMA Extension for Scoping Reviews (Tricco et al., 2018).

1.5. Review question and article organization

Research question: What learner groups, essay-writing tasks, pedagogical strategies, feedback approaches, writing assessment measures and evidence limitations characterize empirical research on Arabic composition among non-native learners published from 2021 through 10 October 2026? The present article develops the review question and an initial evidence chart; comprehensive database exports and independent screening remain outstanding.

Sections 2 and 3 specify the eligibility framework and chart the candidate studies; Section 4 interrogates outcome comparability, Section 5 discloses methodological limitations, and Section 6 states the provisional conclusion.

2. Methods

2.1. Review design and eligibility

The planned review uses Population–Concept–Context (PCC) criteria following JBI guidance (Pollock et al., 2024). Population refers to learners of Arabic as a foreign, second or additional language, including non-native school and university learners.

Concept refers to composing Arabic text, especially paragraphs, expository and argumentative essays, descriptive or creative writing, writing instruction and revision, and assessment of production. Context comprises formal or supported Arabic learning worldwide. 'Arabic essay' denotes sustained written expression and organization, not merely the use of Arabic script; adjacent studies of orthography or morphological form are charted as related evidence, not interpreted as essay-outcome trials.

Eligibility is planned for original peer-reviewed empirical journal articles dated 1 January 2021 through 10 October 2026. Eligible designs include instructional contrasts, classroom action research, qualitative learning-process studies, analyses of learner-produced texts and implemented development projects. Pure reviews, editorials, conference papers, computational systems without learner involvement, studies exclusively of English composition by Arabic-speaking learners, and purely native-language school literacy without distinguishable additional-language data will be excluded. Borderline cases with uncertain learner-language status or no direct writing production are flagged for full-text decision.

2.2. Sources and Boolean strategies

The intended databases are Scopus, Web of Science Core Collection and ERIC, supplemented by multidisciplinary publisher repositories, the Ijaz Arabi, Arabiyat and other journal sites, and relevant regional Arabic-language indexes. Search fields will be adapted to each platform. No institutional subscription search export was available for this draft; targeted journal-site discovery identified candidate studies, and the final search yield is not known.

Core query for title, abstract or keyword fields: (Arabic OR "Arabic as a foreign language" OR "Arabic as a second language" OR "تعليم العربية" OR "اللغة العربية") AND ("essay writing" OR "written composition" OR "paragraph writing" OR "creative writing" OR "writing skill*" OR kitabah OR "مهاره الكتابة" OR "التعبير الكتابي" OR "كتابة المقالة") AND (learn* OR student* OR teach* OR classroom OR pedagog* OR assess* OR feedback OR "غير الناطقين" OR "الناطقين بغيرها").

A precision query for extended composition will be run independently: (Arabic AND ("essay" OR "composition" OR "academic writing" OR "كتابة المقالة" OR "الإشياء")) AND (non-native OR "foreign language" OR "second language" OR undergraduate OR university OR "الناطقين بغيرها"). A sensitivity query will additionally combine orthography, imla, drafting, revision, corrective feedback, and كتابة with Arabic-

learning terms. Platform-specific syntax, Unicode normalization, fields and truncation must be recorded in a dated search log.

Protocol registration and screening accountability: no protocol registration identifier or dated database export is available in the source manuscript. The search equations are proposed strategies, not evidence of an executed search. Accordingly, the twenty records in Tables 1–2 are an exploratory discovery set; actual inclusion status, duplicate counts, reasons for exclusion, and total eligible reports remain unknown.

2.3. Selection, charting and synthesis

The protocol requires two reviewers to independently screen records and full texts, reconcile disagreements, remove duplicate DOIs and maintain reasons for exclusion. Data charting will capture author, year, place, learners' first languages where stated, Modern Standard Arabic versus spoken varieties, educational level, task type, essay genre, assignment length, time on task, pedagogical sequence, comparator, rubric domains, inter-rater reliability, exact quantitative results, delayed test and limitations. A five-paper calibration set should precede full charting.

The present exploratory register comprises twenty original research records whose bibliographic identity and relevant summaries were located through journal publisher pages or recognized national scholarly indexes. These records were not selected by a completed database-level PRISMA pathway and do not establish the number of eligible studies. In particular, some records concern orthography or curriculum design and one university essay study does not explicitly establish non-native learner status in its abstract. Such ambiguity is preserved rather than silently resolved.

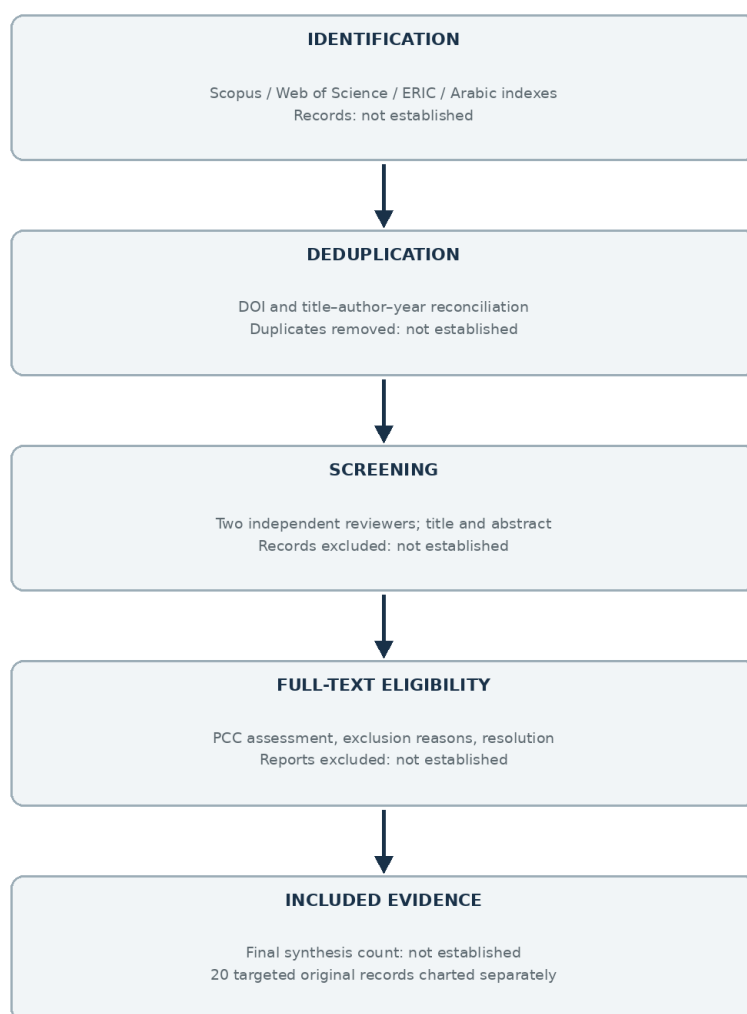
Narrative evidence synthesis is organized by evidence object and design rather than by a league table of favored instructional technologies. Extracted statistical quantities are described as reported values, not recalculated effect sizes. For this draft, source checking establishes article identity and the source's abstract-level claim, with selected publisher full-text snippets where available; independent duplicate extraction of complete research reports has not been accomplished.

2.4. PRISMA-ScR accountability

Figure 1 preserves the intended identification–deduplication–screening–full-text–inclusion sequence. It explicitly leaves every numerical box without a number because export and screening stages have not occurred. Twenty manually charted

candidates cannot be relabeled as 'included after PRISMA'; conflating these steps would create a fabricated study-selection history. PRISMA-ScR is used here as a reporting target rather than as proof that the review has already met all checklist items (Tricco et al., 2018).

PRISMA-ScR workflow: audit fields still pending



Unexecuted database stages have no invented counts.
Diagram is a procedural template, not a completed review flow.

Figure 1. Proposed PRISMA-ScR workflow. Counts are intentionally unfilled; this is not a completed selection diagram.

3. Preliminary Results and Evidence Map

The current chart contains twenty original studies spread across essay composition, paragraph instruction, process writing, feedback, digital supports, orthography and assessment. Figure 2 defines how this body of evidence will be organized, while

Table 1 preserves one row for each located original source. The task unit is central: six studies emphasize essay or paragraph writing, while others treat discrete errors, motivational processes, curriculum or writing resources. These are descriptive counts of the twenty manually located records and have no claim to represent field-wide publication frequencies.

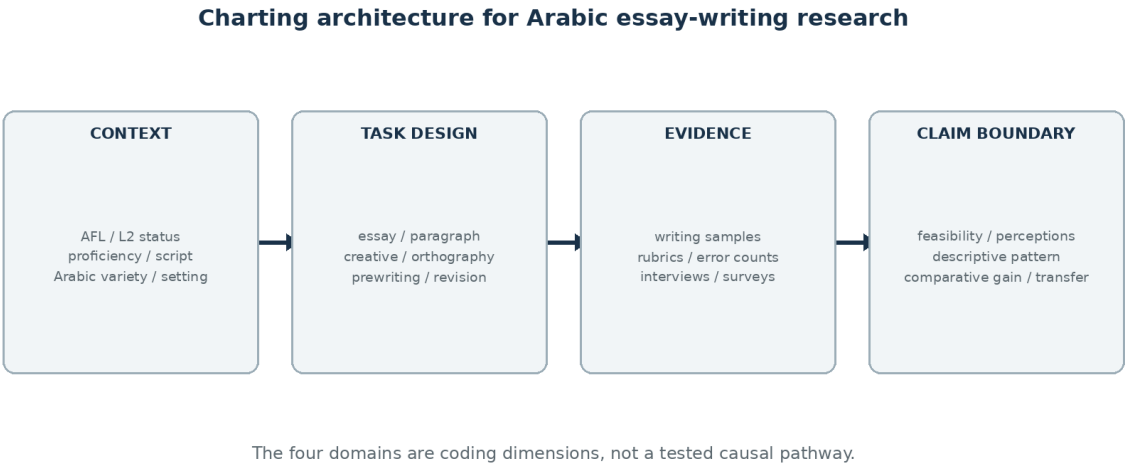


Figure 2. Author-developed charting model. Distinguishing context, task, evidence and claim boundaries prevents inappropriate comparison across designs.

Across the chart, direct writing-production research and indirect writing-related research coexist. The text analysis based on CAF dimensions examines learner-produced language (Syaifullah et al., 2026), whereas the survey on collaborative strategic reading chiefly records learner perceptions (Azman and Bin Chik, 2024). A CEFR-based curriculum project describes possible sequencing of beginner learning aims but does not measure completed essays (Dianto et al., 2022). These examples show why extraction must document the unit of inference, not simply assign all articles to the 'writing' category.

The provisional categories in Table 1 can be read as complementary lenses, not as mutually equivalent quality scores. Essay studies emphasize extended texts; paragraph studies address a smaller rhetorical unit; process and feedback studies capture instructional actions; orthographic studies typically assess local form. The cross-cutting category of assessment draws attention to rubrics, objective scoring and students' actual written products. This taxonomy guards against a common synthesis error: treating an improved response on a spelling test as evidence of better paragraph coherence or argumentative organization.

Studies variously assess learning, implementation and resources: surveys, classroom comparisons, qualitative revision records and instructional development do not answer the same question. Subsequent charting will distinguish comparators, baseline equivalence, rubrics, independent marking and follow-up before interpreting reported gains.

Table 1. Characteristics of original candidate studies

Study	Population / context	Writing focus	Study design	Measured or described evidence
S01 Almugim (2023)	University; learner L1/L2 status needs confirmation	Essay; five-stage process writing	Instructional comparison; details to recheck	Holistic essay performance
S02 Sathikulameen (2023)	India; 25 undergraduate AFL learners	Activity-based paragraph writing	One-group classroom study; questionnaires before/after	Paragraph writing and learner reports
S03 Azman (2024)	Malaysia; 25 upper-secondary learners	Collaborative strategic reading	Cross-sectional questionnaire	Interest, perceived vocabulary and composition support
S04 Amaliyah (2025)	Indonesia; two intact classes (33/33)	Think–Pair–Share; sentences and paragraphs	Quasi-experiment; nonequivalent groups	Posttest writing performance
S05 Kharashi (2025)	Malaysia; 30 intermediate foreign learners	Whole-language creative-writing unit	Quasi-experimental pre/post; single study group	Assessed creative-writing subskills
S06 Arifin (2025)	Indonesia; 122 university students	TPACK, cultural texts and posters	Mixed methods: tests, interviews, observation	Phrases, sentences and descriptive paragraphs

Study	Population / context	Writing focus	Study design	Measured or described evidence
S07 Anwar (2025)	Indonesia; distance-learning Islamic studies students	Online discussion and writing	Qualitative interviews and content analysis	Reported writing experiences and responses
S08 Syaifullah (2026)	Indonesia; AFL university students	Complexity–accuracy–fluency	Descriptive qualitative texts and questionnaire	Syntactic complexity, accuracy, fluency, coherent ideas
S09 Al-Adwan (2025)	Jordan; 37 advanced-level non-native learners	Essay error patterns	Descriptive error analysis of essay test	Spelling and grammatical error codes
S10 Moon (2025)	South Korea; 32 beginner Arabic learners	Teacher feedback differentiated by interest	Four-group pre/post study	Grammar, content, organization, confidence
S11 Deviana (2022)	Indonesia; 65 university Arabic-education students	Writing processes and self-efficacy	Qualitative participant observation	Strategies and learner confidence
S12 Maimuna (2022)	Indonesia; university writing program	Genre-based writing handbook	ADDIE development; validation and pre/post pilot	Text-based writing performance and

Study	Population / context	Writing focus	Study design	Measured or described evidence
				expert validation
S13 Pimada (2021)	Indonesia; entry-level Arabic students	Hamzah spelling error classification	Qualitative observation and writing documentation	Hamzah orthography errors
S14 Zaid (2025)	Indonesia; boarding-school students	Hamzah errors in weekly compositions	Qualitative document analysis	Hamzah washl/qath'i usage
S15 Sadiyah (2026)	Indonesia; 36 university Arabic beginners	Peer editing and error correction	Two-cycle classroom action research	Numeral–noun agreement and mastery
S16 Hamaz (2026)	Indonesia; two intact classes (13/13)	Peer versus teacher feedback in dictation	Quasi-experimental pre/post comparison	Orthographic errors; learner attitudes
S17 Miolo (2025)	Indonesia; 60 high-school students	Kahoot-supported orthography materials	ADDIE development and pre/post trial	Knowledge of writing rules
S18 Al-Wesabi (2026)	Indonesia; 14 male upper-secondary learners	Task-based scaffolding and collaborative revision	Qualitative classroom process within R&D	Peer interaction, revisions, feedback uptake

Study	Population / context	Writing focus	Study design	Measured or described evidence
S19 Dianto (2022)	Indonesia; CEFR A1–A2 learner curriculum	Progression of beginner writing outcomes	Curriculum design study	Task and objective alignment
S20 Khamaiseh (2024)	Malaysia; 123 university learners	Ten morphological verb patterns	Survey and recognition/extraction test	Morphological identification

Candidate chart only. Inclusion status and full-text details remain subject to independent eligibility screening.

Table 2. Claim, measurement and interpretation boundaries

Study	Map strand	Supported statement from publisher/index summary	Interpretive constraint
S01 Almugim (2023)	Essay	Essay-writing performance reported better after a five-stage process approach.	Confirm non-native status; rubric detail unresolved.
S02 Sathikulameen (2023)	Paragraph	Activity-based teaching associated with reported paragraph-writing progress.	Questionnaire emphasis; no robust comparator.
S03 Azman (2024)	Essay	Twenty-five students expressed positive views of reading-to-writing strategy.	Perceived support is not measured essay gain.
S04 Amaliyah (2025)	Paragraph	Posttest means 77.66 versus 58.81; experimental N-gain 0.55.	Intact classes; baseline equivalence requires checking.

Study	Map strand	Supported statement from publisher/index summary	Interpretive constraint
S05 Kharashi (2025)	Essay	Thirty learners showed pre/post creative-writing improvement.	Single-group design limits attribution.
S06 Arifin (2025)	Paragraph	Descriptive writing tasks examined with 122 university students.	No clear independent counterfactual effect.
S07 Anwar (2025)	Digital	Online discussion yielded heterogeneous responses and writing development.	Qualitative accounts cannot establish proficiency gains.
S08 Syaifullah (2026)	Assessment	Writing texts showed complexity, accuracy and idea-development challenges.	Descriptive performance, not treatment effect.
S09 Al-Adwan (2025)	Assessment	Thirty-seven non-native students completed grammar and spelling essay tests.	Error frequencies are not instructional effects.
S10 Moon (2025)	Feedback	Thirty-two beginners assessed across feedback, writing and confidence conditions.	Feedback interaction needs design-level validation.
S11 Deviana (2022)	Process	Observation describes writing instruction oriented to self-efficacy.	Observation does not estimate causal gain.
S12 Maimuna (2022)	Process	Genre handbook validated; pre/post writing gains reported.	Development and pre/post testing lack randomization.
S13 Pimada (2021)	Orthography	Qualitative documentation identified multiple hamzah error categories.	Spelling errors do not index essay coherence.

Study	Map strand	Supported statement from publisher/index summary	Interpretive constraint
S14 Zaid (2025)	Orthography	Weekly compositions showed hamzah washl and qath'i errors.	Diagnosis does not quantify intervention effects.
S15 Sadiyah (2026)	Feedback	Classroom mastery rose from 12/36 to 32/36 across editing cycles.	No independent comparison group across cycles.
S16 Hamaz (2026)	Feedback	Thirteen-versus-thirteen learners compared peer and teacher dictation feedback.	Dictation outcome differs from essay production.
S17 Miolo (2025)	Digital	Orthography scores reportedly rose from 58.85 to 84.20.	Writing-rules knowledge is not essay quality.
S18 Al-Wesabi (2026)	Process	Fourteen learners' task-based scaffolding and revisions documented.	Process evidence does not estimate performance gain.
S19 Dianto (2022)	Assessment	Curriculum design proposes sequenced CEFR A1–A2 learning objectives.	Curriculum proposal is not learner outcome data.
S20 Khamaiseh (2024)	Assessment	Ten verb-pattern recognition examined among 123 learners.	Pattern recognition is not composition quality.

The evidence statements are deliberately conservative. Some records are adjacent-domain candidates, not verified final inclusions.

3.1. Process writing, genre and creative composition

A significant strength of essay-focused instruction is that its target outcome can be specified as a full text rather than as correctness on isolated items. Almugim's writing-process approach organizes preparation and drafting into successive stages and reports better undergraduate essay performance than conventional practice (Almugim, 2023). Genre-oriented instructional materials incorporate building field knowledge, modeling, joint construction and independent construction (Maimuna et al., 2022). Both approaches treat writing as sequenced text production rather than a one-time recall exercise. Nevertheless, their studies differ in materials, population definition and methodological description, and the available summaries cannot isolate which instructional stage produced the reported change.

Other instructional studies support a broader composition focus. The whole-language unit evaluated with thirty intermediate foreign learners reports improved creative-writing subskills after instruction (Kharashi and Mohd Noor, 2025). Activity-based writing work conducted with twenty-five undergraduates in India reports indications of paragraph-writing progress, but its abstract describes questionnaires before and after activities instead of a fully described composition-scoring design (Sathikulameen et al., 2023). In both cases the appropriate analytic questions concern rubric structure, control conditions and whether the assigned genre permits inference about independent essay writing.

3.2. Interaction, peer feedback and digital mediation

The Think–Pair–Share study provides an explicit nonequivalent-control-group design and gives between-group posttest means for Arabic writing (Amaliyah et al., 2025). These statistics are potentially useful, but intact classes can differ before treatment. By contrast, observations within a fourteen-learner task-based program trace how linguistic scaffolding, dialogue, feedback and revision interact (Al-Wesabi et al., 2026). This process evidence helps articulate plausible writing mechanisms but does not determine the size of learning effects.

Two newer feedback studies address narrower accuracy targets. Sa'diyah et al. describe sequential gains in numeral–noun agreement accuracy through peer editing (Sa'diyah et al., 2026), and a separate comparison of two thirteen-student classes examines peer versus teacher feedback for dictation errors (Hamaz et al., 2026). Their findings concern local orthographic or morphosyntactic accuracy, which may be necessary for extended writing yet cannot independently establish improved essay argumentation, coherence or creative expression.

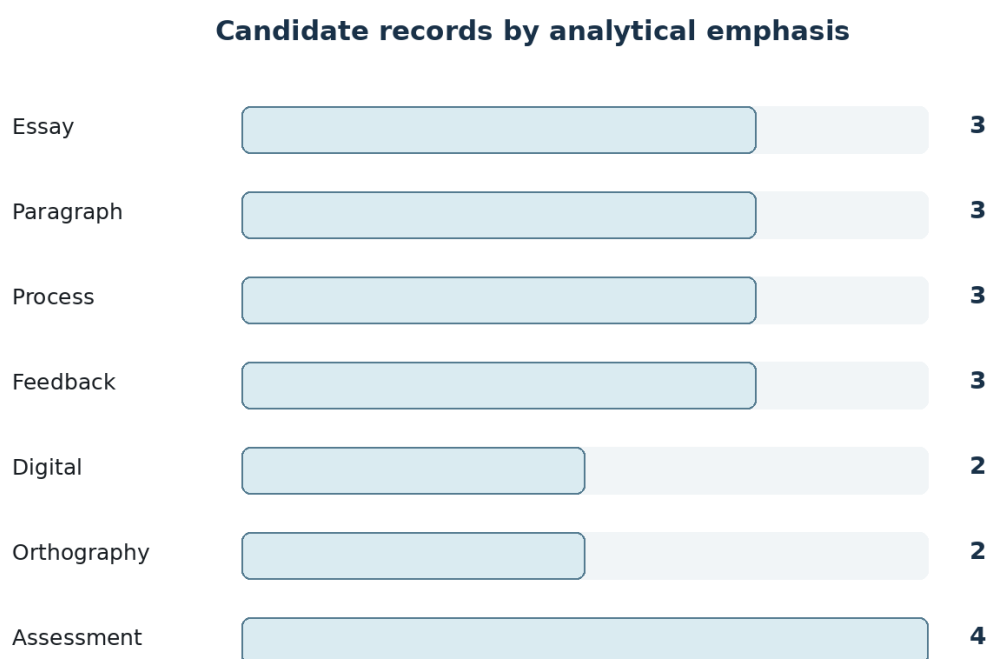
The pedagogical promise of technology appears in several formats that should not be conflated. Poster tasks structured around TPACK and multicultural description involved 122 university students and yielded reported changes in written phrases, sentences and paragraphs (Arifin et al., 2025). Online discussion, meanwhile, produced diverse learner experiences and mixed perceived writing development (Anwar et al., 2025). An ADDIE-designed Kahoot orthography resource produced a reported pre/post shift from 58.85 to 84.20 on writing-rules learning (Miolo et al., 2025). The first examines descriptive text, the second participation and experience, and the third orthography knowledge; none is a direct comparison of independently written argumentative essays against a common rubric.

3.3. Linguistic diagnosis and assessment

Error analysis supplies precise information about local constraints on writing. One study used an essay test with thirty-seven non-native Arabic students in Jordan to code grammatical and spelling errors (Al-Adwan and El-Shara, 2025). A separate study analyzed recurrent hamzah washl and hamzah qath'i errors in weekly compositions (Zaid et al., 2025), building on earlier qualitative documentation of hamzah difficulties among beginner learners (Pimada et al., 2021). These studies identify instruction-relevant problem types but do not, by themselves, demonstrate the outcomes of any corrective intervention.

Broader outcome frameworks need to be disaggregated. The CAF study evaluates complexity, accuracy and fluency alongside students' reported affective experiences (Syaifullah et al., 2026). A survey-and-test investigation of ten Arabic morphological verb forms involved 123 learners, yet the publisher abstract emphasizes recognition and extraction of patterns rather than an evaluated essay text (Khamaiseh et al., 2024). Treating its results as a direct measure of argumentative or expository writing would overstate the outcome's scope.

Assessment of writing also includes instructional context. A Korean four-group study with thirty-two beginner learners compared teacher feedback types under varying degrees of interest and measured writing grammar, content, organization and confidence (Moon, 2025). Self-efficacy-oriented classroom observation identifies how writing-related strategies are enacted in a university cohort (Deviana et al., 2022). These studies suggest that writing measures need both linguistic and higher-order components, but their heterogeneous designs prevent any present estimate of a universal method effect.



Descriptive counts within 20 targeted candidate records only; not PRISMA estimates.

Figure 3. Classification counts within the twenty targeted candidate records. Categories are mutually assigned for display purposes; they do not estimate the full research field.

4. Discussion

4.1. A review of writing objects rather than method labels

The most useful insight of the candidate evidence is that 'Arabic writing' refers to multiple targets. The outcome of dictation feedback is often an error rate; the outcome of a process-writing intervention may be a complete text; and the outcome of collaborative strategic reading may be a questionnaire about confidence or idea selection. Table 1 clarifies the empirical unit, and Table 2 tests the boundary between what each publication reports and what can legitimately be inferred. A narrative synthesis that grouped these studies only by technological or teaching method would ignore the central variable: the textual object actually assessed.

Figure 2 provides the framework for correcting this ambiguity in the completed review. Context and learner background enter first, followed by the writing task and evidence captured. The final analytic category is not 'success' versus 'failure' but the claim's permissible strength. A qualitative description of interaction can be strong evidence about the process of collaboration while remaining unable to establish a treatment effect. Conversely, a short pre/post test can estimate

immediate task performance but cannot establish durable ability to produce an independent essay.

Measurement crosswalk: a complete essay rubric should separate content development, rhetorical organization, coherence and cohesion, linguistic accuracy, and task fulfillment. A hamzah or dictation score captures only part of this construct. Where texts have multiple raters, the completed review should chart rater training and agreement; where text is digitally assisted, it should distinguish guided revisions from independent writing.

4.2. The difference between orthographic accuracy and writing quality

Arabic-language education includes substantial attention to hamzah, imla' and morphological patterns. Their instructional relevance is clear: errors may disrupt readability or draw learner attention away from discourse construction. However, composition is a larger construct. A study can report fewer hamzah errors without investigating whether learners developed thesis statements, arguments, transitions or conclusions. The contrast between text-based error analysis (Zaid et al., 2025) and holistic composition instruction (Almugim, 2023) therefore offers a reason to chart 'surface' and 'deep' writing outcomes separately.

Evidence from cross-language research strengthens the measurement rationale without settling Arabic-specific questions. In a meta-analysis spanning 200 comparisons, Scherer and colleagues distinguished feedback on surface and deep outcomes and reported different effects by learner population (Scherer et al., 2024). That synthesis does not license applying its effect sizes directly to Arabic, given differences in orthographic and morphological systems. It does, however, support the conceptual demand for matching feedback content to the dimension of writing actually tested.

4.3. Process writing and opportunities for transfer

The process-based and genre-based strands are more closely aligned with essay production than tools that merely rehearse spelling rules. Their shared characteristics include activating background knowledge, modeling textual organization, supporting planning, staging drafts and enabling revision. The handbook development work distinguishes modeled from independent construction (Maimuna et al., 2022), and the university essay study includes explicit prewriting preparation (Almugim, 2023). The critical unresolved issue is whether improvements persist when scaffolds and model texts are removed and learners face new essay prompts.

Genre-specific outcomes can also vary by context. Creative writing requires elaboration, audience engagement and narrative or descriptive control; academic essays may require claims, warrants, organization and citation. The whole-language creative-writing study (Kharashi and Mohd Noor, 2025) and multicultural poster-writing study (Arifin et al., 2025) therefore cannot be assumed to measure the same construct. A defensible completed map will code genre, prompt complexity, length and scoring domains before determining which studies are comparable.

4.4. Feedback ecology and learner agency

Feedback is not a unitary treatment. The peer-editing action study focuses on correcting a specific grammatical relationship (Sa'diyah et al., 2026). The two-class dictation study separates peer and teacher feedback (Hamaz et al., 2026), while qualitative task-based research records noticing, interaction and revision as classroom processes (Al-Wesabi et al., 2026). Combining these distinct designs may be useful to understand a feedback ecology, but not for calculating an overall 'peer feedback works' conclusion. Intervention duration, opportunities to revise, teacher input and task complexity need explicit charting.

Learner interest and confidence can be important implementation outcomes without serving as proof of essay proficiency. The Malaysian survey of collaborative reading measured what students believe helps them select and express ideas (Azman and Bin Chik, 2024). Korean feedback research includes confidence alongside scored writing dimensions (Moon, 2025). Preserving these distinctions makes it possible for future synthesis to relate learner experience to performance while not treating them as interchangeable.

Independent authorship and transfer are especially important for any AI-assisted or digitally mediated writing task. The current candidate records do not establish comparable conditions for assisted and unassisted essays. Thus neither improved accuracy under tool guidance nor positive attitudes to a platform can substantiate durable essay-writing competence without an appropriately defined writing product.

4.5. Digital tools and evidence of independent authorship

Digital resources may provide useful writing prompts, vocabulary access, models and feedback, yet their capabilities differ. Poster media foreground cultural description (Arifin et al., 2025), online discussion makes written interaction visible (Anwar et al., 2025), and Kahoot material targets orthographic rules (Miolo et al., 2025). Outcome claims therefore ought to identify the exact activity and which parts

of a text were authored independently. Especially in emerging AI-supported essay writing, future designs will need to distinguish assisted correction from unassisted performance and report the degree of learner control over final text.

Figure 3 shows that the current manually assembled corpus spans seven analytic strands. The numerical balance reflects targeted discovery only; it cannot be used to assert that assessment research dominates the worldwide field. Its main value is practical: it shows which coding fields are needed before a larger reproducible search and which adjacent domains must be screened carefully rather than automatically admitted.

4.6. Evidence gaps and readiness for further synthesis

The candidate studies indicate several prospective gaps, not yet field-wide prevalence claims. First, learner backgrounds, Arabic exposure, proficiency benchmarks and target Arabic variety are incompletely visible in publisher abstracts. Second, some writing studies rely heavily on pre/post contrasts without clear independent comparison groups. Third, immediate writing scores are reported more readily than delayed, novel-prompt or transfer measures in the small targeted set. Fourth, qualitative accounts frequently illuminate process without comparable writing samples or transparent scoring rubrics.

A completed map could test these observations against all eligible studies and identify which writing genres support focused systematic reviews. For example, a sufficiently large cluster of comparable rubric-based composition experiments might eventually justify a narrower effectiveness synthesis. At present, pooling the reported means from classroom projects, corpus analyses, spelling tests and perception questionnaires would violate outcome comparability and produce a misleading estimate.

Cross-study count sensitivity: Figure 3 partitions these twenty manually identified candidates into seven working categories (essay 3, paragraph 3, process 3, feedback 3, digital 2, orthography 2, and assessment 4). These numbers reflect analyst coding of a convenience discovery set; they must not be interpreted as database prevalence, an exhaustive literature census, or a completed PRISMA study count.

5. Strengths, Limitations and Research Integrity

The manuscript's conceptual strengths are its restricted focus on Arabic writing, explicit recognition of differing textual targets, international retrieval terms in more than one language and a citation register tied to publisher-linked records. The principal limitation is fundamental: the research has not yet completed independent

Scopus, Web of Science and ERIC searches, full-text screening, deduplication or PRISMA reconciliation. The twenty original studies are a targeted candidate set, not an exhaustive or representative synthesis. Their distribution across teaching approaches and countries cannot support prevalence claims.

Several entries have only publisher abstract or bibliographic-level verification, and some have unresolved eligibility questions. The table purposely retains these uncertainties. The final study would require stable search dates, reproducible exports, duplicate review decisions, cross-language full-text assessment, exact author metadata, independently checked numeric outcomes and optional critical appraisal tailored to study design. No arbitrary 40–60-study target or 100–150-citation target should override eligibility and source accuracy.

6. Conclusion

Research on Arabic essay writing among non-native learners is best approached as a set of distinct writing and learning objects rather than a contest among pedagogical labels. The exploratory evidence suggests promising avenues in process instruction, task-based collaboration, feedback and genre scaffolding while simultaneously documenting the importance of distinguishing essay organization from orthographic accuracy and learner perceptions. The proposed charting model makes these dimensions auditable. Only a completed multilingual, independently screened PRISMA-ScR review can establish the full extent and distribution of evidence or make defensible field-wide claims. Until then, the findings are hypotheses for systematic mapping, not publication-ready conclusions.

References

- Al-Adwan, H. M. M., & El-Shara, I. A. H. (2025). Writing and grammatical errors among non-native Arabic language learners. *Jordanian Educational Journal*, 10(3), 93–118. <https://doi.org/10.46515/jaes.v10i3.1564>
- Al-Wesabi, A. A. M., Solahudin, D., Syafei, I., & Kosim, N. (2026). Scaffolding, collaboration, and feedback in task-based Arabic writing: Evidence from an Indonesian Islamic boarding school. *Al-Muyassar: Journal of Arabic Education*, 5(3), 851–868. <https://doi.org/10.31000/g972ey10>
- Almugim, T. M. (2023). The effect of using the writing processes approach on improving the essay writing skill of undergraduate students. *Ijaz Arabi Journal of Arabic Learning*, 6(1). <https://doi.org/10.18860/ijazarabi.v6i1.20436>

- Amaliyah, P. K., Maswani, M., & El Zahraa, F. (2025). The effectiveness of using the Think-Pair-Share strategy in teaching Arabic writing skills. *Al-Muyassar: Journal of Arabic Education*, 4(2), 321–344. <https://doi.org/10.31000/al-muyassar.v4i2.13545>
- Anwar, M. S., Ulum, M. S., Maghfiroh, R., Hamidah, N. Z., & Munir, M. S. (2025). Using an online discussion platform to teach writing skills for Islamic studies students. *Ijaz Arabi Journal of Arabic Learning*, 8(1). <https://doi.org/10.18860/ijazarabi.v8i1.23695>
- Arifin, Z., Febriani, S. R., Mahmudi, A., & Bedra, K. G. (2025). Optimizing Arabic writing skills based on TPACK learning and poster media. *Ijaz Arabi Journal of Arabic Learning*, 8(3). <https://doi.org/10.18860/ijazarabi.v8i3.32329>
- Azman, Y., & Bin Chik, A. R. (2024). Students' views on using collaborative strategic reading (CSR) to develop Arabic composition writing skills. *Ijaz Arabi Journal of Arabic Learning*, 7(2). <https://doi.org/10.18860/ijazarabi.v7i2.26392>
- Deviana, A. D., Qomariyah, L., Mubarak, F., & Ahya, A. S. (2022). Learning strategies for writing skills with self-efficacy based. *Ijaz Arabi Journal of Arabic Learning*, 5(1). <https://doi.org/10.18860/ijazarabi.v5i1.13196>
- Dianto, T. N., Joko P, N., Fatoni, A., & Kalita, S. (2022). CEFR-based beginner Arabic reading and writing curriculum design in Indonesia. *Ijaz Arabi Journal of Arabic Learning*, 5(3). <https://doi.org/10.18860/ijazarabi.v5i3.16684>
- Hamaz, N. L., Lestari, O. D., Subhi, M., & Musthofa, T. (2026). The effect of peer feedback and teacher feedback on the improvement of imla' errors and the psychological attitudes of eighth-grade students at MTs Al-Fath Klaten. *Abjadia: International Journal of Education*, 11(1), 203–211. <https://doi.org/10.18860/abj.v11i1.39881>
- Khamaiseh, M. A., Osman, N., & Elawadi, E. M. S. (2024). Morphological errors in the ten wazan for the writing skill of non-native Arabic language learners. *Ijaz Arabi Journal of Arabic Learning*, 7(1). <https://doi.org/10.18860/ijazarabi.v7i1.26073>
- Kharashi, N. A. R., & Mohd Noor, S. S. (2025). Evaluating the effectiveness of an educational unit using the whole language approach to develop Arabic creative writing skills for intermediate level. *Ijaz Arabi Journal of Arabic Learning*, 8(1). <https://doi.org/10.18860/ijazarabi.v8i1.31198>

- Maimuna, M., Ainin, M., & Taufiqurrahman, R. (2022). Arrangement of text-based handbooks to enhance writing skills in Arabic language education. *Ijaz Arabi Journal of Arabic Learning*, 5(1). <https://doi.org/10.18860/ijazarabi.v5i1.15992>
- Miolo, M. I., Hula, I. R. N., & Laiya, F. F. S. (2025). Arabic orthography: Development of Kahoot and gamification website-based writing materials to improve understanding of writing rules. *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 12(1), 1–17. <https://doi.org/10.15408/a.v12i1.47225>
- Moon, J.-Y. (2025). Effects of teacher feedback types on beginner learners' writing skills and confidence based on their interest level in Arabic. *Arabic Language and Literature [Korean journal]*, 29(1), 55–84. <https://doi.org/10.18630/kaall.2025.29.1.003>
- 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>
- Pimada, L. H., Amrulloh, M. A., Zahra, D. N., & Syahid, A. H. (2021). The errors in writing hamzah: What should students do?. *Ijaz Arabi Journal of Arabic Learning*, 4(1). <https://doi.org/10.18860/ijazarabi.v4i1.9870>
- 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 *JBIM Manual for Evidence Synthesis* (2024 ed.). JBI. <https://doi.org/10.46658/JBIMES-24-09>
- Sa'diyah, H., Rusuly, U., & Saputri, I. N. F. (2026). Implementing the PEEC technique through classroom action research: A strategy to mitigate adad-ma'dud errors in beginner-level Arabic writing. *Asatiza: Jurnal Pendidikan*, 7(1), 78–92. <https://doi.org/10.46963/asatiza.v7i1.3390>
- Sathikulameen, A., Ahamed Zubair, K. M. A., Nadwi, A. A. H. H., Ahamedullah, M., & Ibrahim, A. M. A. (2023). Effect of an activity-based learning in teaching Arabic paragraph writing to AFL learners. *Ijaz Arabi Journal of Arabic Learning*, 6(3). <https://doi.org/10.18860/ijazarabi.v6i3.22939>
- Scherer, S., Graham, S., & Busse, V. (2024). How effective is feedback for L1, L2, and FL learners' writing? A meta-analysis. *Learning and Instruction*, 93, Article 101961. <https://doi.org/10.1016/j.learninstruc.2024.101961>

- Syaifullah, S., Jatmiko, R. S., Fransisca, M., Mahendra, A., Kurniawati, A., & Setyawan, C. E. (2026). CAF dimensions as measures of university students' writing proficiency in Arabic as a foreign language. *Ijaz Arabi Journal of Arabic Learning*, 9(2). <https://doi.org/10.18860/ijazarabi.v9i2.37691>
- 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>
- Zaid, A. H., Putri, S. F., & Anggraini, R. M. (2025). Analysis of nahwu errors in hamzah wasl and hamzah qath'i class 5 KMI at Gontor for Girls. *Ijaz Arabi Journal of Arabic Learning*, 8(1). <https://doi.org/10.18860/ijazarabi.v8i1.29119>