



Teaching and Assessing Arabic Listening Comprehension among Learners of Arabic as a Foreign or Second Language: A Scoping Review of Instructional Approaches, Digital Media, and Evidence Gaps (2021–2026)

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 Thoriquattyas

Universitas Negeri Malang, Indonesia

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

Abstract

Arabic listening comprehension entails auditory segmentation, lexical recognition and discourse interpretation under variable speech conditions. This scoping-review project maps instruction, assessment and research gaps among learners of Arabic as a foreign or second language (2021–2026). Using a Population–Concept–Context framework, an exploratory search of journal and publisher sites located 41 candidate original studies for preliminary charting; systematic searches of Scopus, Web of Science and ERIC, independent screening and full-text extraction are not yet complete. Candidate studies address task-based teaching, podcasts, audiovisual media, metacognition, gamification and AI-supported microlearning. Reported outcomes include feasibility scores, perceptions, correlations and within-group listening gains, which do not support a pooled effectiveness claim. Future research needs validated listening tasks, explicit input-register descriptions, appropriate comparators and transfer measures. These findings remain provisional until the PRISMA-ScR study-selection process is completed.

Keywords: Arabic listening comprehension; Arabic as a foreign language; digital pedagogy; scoping review.

مستخلص البحث

تبحث هذه المراجعة الاستكشافية في تدريس مهارة الاستماع باللغة العربية وتقييمها لدى متعلمي العربية بوصفها لغة ثانية أو أجنبية خلال المدة من ٢٠٢١ إلى ٢٠٢٦. اعتمد العمل إطار السكان والمفهوم والسياق ومبادئ الإبلاغ في PRISMA-ScR. وقد حدد البحث الاستطلاعي في مواقع الناشرين إحدى وأربعين دراسة أولية مرشحة لرسم خريطة مبدئية للأدلة، ولم يكتمل بعد البحث المنهجي في قواعد البيانات أو فحص النصوص الكاملة. وتتناول الدراسات المرشحة التعلم القائم على المهام والمدونات الصوتية والوسائط السمعية البصرية والاستراتيجيات فوق المعرفية والألعاب التعليمية والتعلم المدعوم بالذكاء الاصطناعي. وتختلف المخرجات بين تقييم صلاحية الوسائط وآراء المتعلمين والارتباطات والتغيرات بين الاختبارين القبلي والبعدي. لذلك لا تسمح المعطيات الحالية بالحكم على تفوق طريقة تعليمية معينة. وتوصي المراجعة بتحسين صدق اختبارات الاستماع ووصف نوع العربية المسموعة وتوثيق مجموعات المقارنة واختبار انتقال التعلم واستمراره. وتظل النتائج أولية حتى اكتمال إجراءات البحث والاختيار والتحقق.

الكلمات المفتاحية: فهم المسموع؛ تعليم العربية للناطقين بغيرها؛ الوسائط الرقمية؛ المراجعة الاستكشافية.

1. Introduction

1.1. Background and importance of Arabic listening

Listening comprehension is central to learning Arabic in environments where students encounter the language mainly through classroom speech, recordings, and digital resources. A listening task requires more than sound exposure: learners discriminate phonemes, segment speech, identify vocabulary and grammatical cues, infer meaning, and retain propositions while new input arrives. Arabic listening is additionally conditioned by Modern Standard Arabic–dialect variation, unfamiliar prosody, prior script literacy, and the linguistic backgrounds of learners. Hence, phoneme recognition, dialogue gist, and inference from a news report are related but nonidentical constructs.

1.2. Instructional and methodological developments

The instructional repertoire documented in recent Arabic-listening research is expanding. Case studies have mapped stories, songs and conversations (Chamidah & Istiqomah, 2021), and teachers have incorporated audio and images into familiar presentation platforms (Berutu et al., 2022). Development studies examine podcasts and listening textbooks (Rahmah & Ahsanuddin, 2022; Subakir et al., 2023), whereas experimental and classroom-action reports explore digital editing, online listening platforms and task-based sequencing (Fikri et al., 2024; Hassan & Shehata, 2025). Such diversity generates important opportunities for practice but also complicates claims that one tool ‘improves listening’ more than another.

1.3. Evidence and measurement gaps

Listening research also differs in the kinds of evidence it collects. Abdul Rahman and Baharudin (2023) surveyed 300 Malaysian secondary students concerning metacognitive awareness and listening achievement. Hamad and Younes (2024) surveyed 127 Arabic learners in Turkey about pre-, during-, and post-listening strategies. Neither report is an intervention experiment. By contrast, development studies may present expert validation and learner reactions without using an independently scored novel listening task (Rahmah & Ahsanuddin, 2022; Maulana et al., 2025). Such distinctions should shape both eligibility decisions and synthesis.

The rise of video, gamification and AI intensifies the interpretive problem. Theoretical affordances include repeatable audio, captions, multimodal scaffolds, tailored pacing and rapid feedback, but studies must specify whether students actually understand unfamiliar speech without those supports. Hikmawati et al. (2025) examine critical-thinking indicators during Kahoot-supported listening, whereas Sofian et al. (2026) evaluate AI-supported vlog microlearning in a 60-student one-group design. These studies address different questions, and conclusions about long-term receptive competence require assessment designs that the media label alone cannot provide.

1.4. Objectives and contribution

Prior reviews have summarized listening/speaking methods (Wahyuni et al., 2023) and digital media in Arabic listening (Yonan et al., 2026). The latter screened 150 records and included 13 studies from Garuda, DOAJ, and ERIC, according to its publisher abstract; those counts belong to that review, not the present project. The contribution proposed here differs by mapping listening constructs and input conditions alongside instructional mechanisms, outcome validity, and the status of

each evidence claim. Consistent with Arksey and O'Malley (2005), Munn et al. (2018), and Tricco et al. (2018), the task is evidence characterization rather than pooled intervention-effect estimation.

1.5. Article organization

This manuscript contains an explicitly provisional map of 41 candidates and a reproducible protocol to complete the review. Figure 3 introduces the conceptual alignment model, Figure 2 documents the exploratory publication-year distribution, and Tables 1–2 chart the candidate records. A finished scoping review will require a dated database search, deduplication, independent decisions, complete full-text charting and reconciled PRISMA-ScR counts before making field-wide conclusions.

2. Review Questions and Objectives

The overarching question is: What types of educational research have investigated the teaching, learning and assessment of Arabic listening comprehension among learners of Arabic as a foreign or second language from 1 January 2021 to 10 October 2026?

RQ1. Which listening constructs—sound discrimination, lexical recognition, gist, specific information, inference, discourse tracking, listening strategy use and transfer—have been investigated?

RQ2. Which instructional approaches and listening-input modalities (live speech, edited audio, podcast, video, mobile platforms, games or AI support) are described, and how are they implemented?

RQ3. Which populations, national and institutional contexts, study designs, task genres, assessment measures, comparators and follow-up intervals appear in the research?

RQ4. Which recurring design, reporting, equity and assessment-validity gaps limit comparison or inform the agenda for future primary research?

3. Methods

3.1. Design, registration, and reporting standard

The intended study design is a scoping review using Population–Concept–Context (PCC) eligibility criteria and the PRISMA Extension for Scoping Reviews (PRISMA-ScR; Tricco et al., 2018). This preliminary evidence chart was compiled from targeted journal and publisher-site discovery, not a comprehensive, deduplicated bibliographic search. No prospective protocol registration, complete two-reviewer selection process or final eligible-study denominator is claimed. The distinction between preliminary charting and completed review evidence is maintained throughout.

For an executable review, the protocol should be frozen before final screening, with dated search logs, DOI-based deduplication, two reviewers and a preserved eligibility-decision file. Figure 1 outlines the stages but deliberately contains no invented database counts. Its single reported exploratory total (41) denotes manually identified candidate records only and is not an inclusion count.

3.2. Population–Concept–Context and eligibility

Population: non-native learners studying Arabic as an additional, foreign or second language, from primary education through university and structured adult instruction. Studies of Arabic first-language listening are excluded unless non-native learners are separately analyzable.

Concept: instruction, learning processes, explicit listening strategies, digital or audiovisual listening media, listening tasks, listening-related materials development, diagnostic difficulty, assessment or listening outcomes. Listening must be an identifiable task, study outcome or curriculum focus; studies of speaking alone or of Arabic language learning without a discernible listening component are not eligible.

Context: formal and structured non-formal education in any country. Modern Standard Arabic is not assumed if the study explicitly uses colloquial or mixed varieties; speech register, native-speaker status and audio authenticity will be charted wherever reported. Original peer-reviewed quantitative, qualitative, mixed-method, classroom action and research-and-development reports published from 2021 to the 10 October 2026 cut-off may be eligible. Editorials, opinion essays, narrative reviews, systematic reviews, theses, unretrievable records and purely technical automatic-speech-recognition papers without learner listening evidence are excluded from the primary-study set. Methodological scoping-review sources and previous reviews may be cited for background but are not primary records in Tables 1–2.

3.3. Information sources and executable search strategy

Planned sources comprise Scopus, Web of Science Core Collection, ERIC and Arabic-language journal collections. Search strings are proposed below but were not executed through subscription database interfaces for this revision. Individual database searches should be logged with search date, interface, exact Boolean expression, language and study-type filters, exported records, and count, followed by reference-list checking. A draft Scopus expression is:

TITLE-ABS-KEY(("Arabic as a foreign language" OR "Arabic as a second language" OR "non-native Arabic" OR "Arabic language learner*" OR "Arabic learner*" OR "maharah istima*" OR "maharat al-istima*" OR "Arabic listening") AND (listen* OR "auditory comprehension" OR "speech comprehension" OR podcast* OR "audio visual" OR audiovisual OR dictation) AND (teach* OR learn* OR instruct* OR assess* OR strateg* OR intervent* OR classroom OR digital)) AND PUBYEAR > 2020 AND PUBYEAR < 2027.

The planned Web of Science counterpart applies TS=(...) to the equivalent lexical set with publication-year limits; ERIC searches will use title, abstract and descriptors with a separately logged Arabic-language query. Supplementary Arabic terms include “مهاراة الاستماع”, “تعليم الاستماع”, “فهم المسموع”, “تعليم العربية للناطقين بغيرها”, “استراتيجيات”, “الاستماع”, and “الاستماع باللغة العربية”. Sensitive searches must also test transliteration variants and avoid filtering solely on English abstracts.

Execution dates, raw database exports, total retrieval counts, duplicate counts and independent full-text exclusion logs are not available for this manuscript. Thus, the 41 publisher-discovered records are not presented as a PRISMA-selected corpus,

and all Figure 1 screening cells requiring audited numbers remain unfilled. The eventual review should also explicitly document the Arabic-language terms and transliteration variants tested.

3.4. Selection procedure and data charting

At final implementation, records will be deduplicated first by DOI and then by normalized title, author and year. Two reviewers should pilot 20 abstracts to standardize interpretation of 'Arabic listening' and the non-native learner boundary, then screen titles/abstracts and retrieve potentially relevant full texts independently. Disagreements should be resolved by discussion or a third reviewer. Full-text exclusions should receive one primary mutually exclusive reason, e.g., no listening-specific outcome, non-eligible population, not an original study or not within the date range.

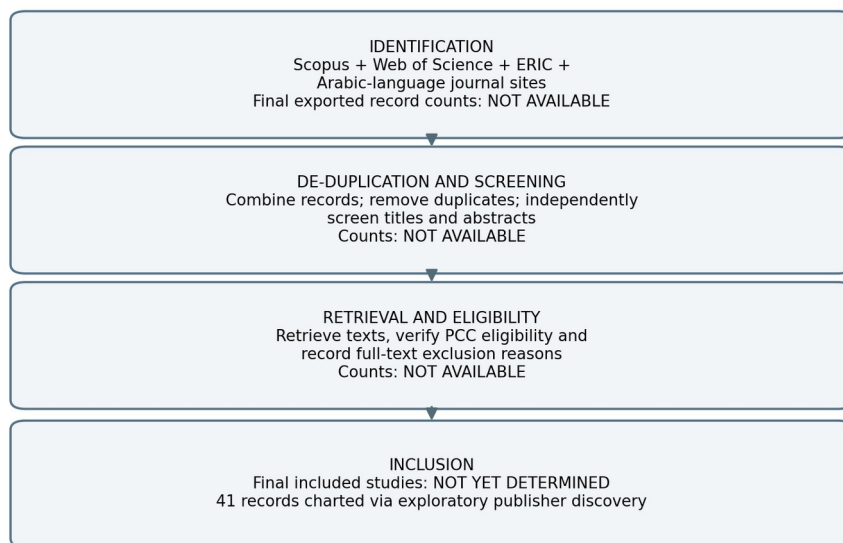
The charting template should include bibliographic identity, country, institution, learner first language and Arabic proficiency, age, educational level, varieties of Arabic, content genre, audio length/rate, task difficulty, modality, pedagogy and comparison group, duration, exposure dosage, sample size, test construction and validation, measurement domains (gist, detail, lexicon, inference, etc.), rater and scorer details, missing-data handling, immediate/delayed performance, retention and transfer, and the study authors' main limitations.

The 41 records were located using exploratory publisher-oriented discovery and have been charted chiefly from available article pages and abstracts. Several key records have since been checked against official publisher pages, but the remaining records have not undergone uniform full-text extraction or database-indexing verification. Unconfirmed details are marked NR rather than estimated.

3.5. Synthesis and claim discipline

The planned analysis is descriptive mapping of study characteristics, then qualitative grouping by teaching modality, listening construct and evidence type. No pooled efficacy comparison will be attempted. The review distinguishes: (i) descriptions of curricula and practices; (ii) perceptions and satisfaction; (iii) validation or feasibility of media; (iv) associations between learner characteristics and listening measures; (v) uncontrolled pre/post change; and (vi) controlled between-group effects where demonstrably reported. This separation prevents the common error of treating product validation percentages as gains in language comprehension.

Critical appraisal is not a prerequisite for every scoping review, but reporting adequacy is central to the present question. We therefore chart whether authors report a credible listening task, construct alignment, allocation/comparator, follow-up, and language-variety specifications without assigning an unvalidated universal study-quality score.



PRISMA-ScR workflow template — not a completed PRISMA flow diagram

Figure 1. Planned PRISMA-ScR workflow. The 41 charted publisher records are preliminary candidates, not final eligible studies; database screening totals remain unreported.

Figure 1 preserves the identification–screening–eligibility–inclusion logic but does not substitute a fabricated numerical flow for a completed search. Once the review is implemented, record counts should be populated directly from the export and reviewer-decision logs and reconciled arithmetically.

4. Provisional Results: Study Characteristics

Forty-one candidate primary-study records dated 2021–2026 were charted provisionally. These are not final PRISMA-ScR inclusions. The working record distribution is 2 (2021), 5 (2022), 6 (2023), 7 (2024), 9 (2025) and 12 (2026). As discovery was targeted rather than comprehensive, these values are descriptive of this working dataset only and cannot demonstrate a publication trend.

Table 1 provides one row per candidate record, presenting educational context, research design, sample and main evidence domain. Table 2 provides one row per candidate record for an abstract-level claim and the most consequential interpretive caveat. Full-text eligibility could remove or reclassify records, particularly those whose listening outcome is embedded in speaking instruction or whose media-development paper reports only feasibility.

The mapped research represents elementary schools, secondary education, boarding schools, universities and preparatory language programs, including examples from Indonesia, Malaysia and Turkey. Such distribution is a property of accessible records, not the global scholarly literature. Arabic-language journal retrieval and further database coverage are likely to alter institutional and geographic patterns.

Table 1. Exploratory charting of candidate primary studies: bibliographic identity, context and design (n = 41; no final PRISMA inclusion decisions).

ID	Study	Context	Design (reported)	Sample*	Primary domain
S01	Rochma et al. (2021)	Indonesian university	qualitative implementation study	NR	Authentic online audio
S02	Chamidah & Istiqomah (2021)	Indonesian Arabic course	qualitative case study	NR	Collaborative/communicative
S03	Berutu et al. (2022)	Indonesian classrooms	descriptive qualitative	NR	Digital media
S04	Rahmah & Ahsanuddin (2022)	Indonesian university	ADDIE product development	20	Podcast/material development
S05	Fawzani et al. (2022)	Indonesian university	cross-sectional correlational	68	Music and audio exposure
S06	Kusumadewi et al. (2022)	Indonesian elementary school	reported quasi-experimental comparison	10	Audiovisual intervention
S07	Mahbub & Khusnul (2022)	Indonesian secondary school	qualitative development account	NR	Audiovisual materials
S08	Subakir et al. (2023)	Indonesian Islamic boarding school	R&D with learner testing	33 + 60 trials	Curriculum/material development
S09	Rosyad et al. (2023)	AFL learners / material design	descriptive design study	NR	Authentic/contextual input
S10	Abdul Rahman & Baharudin (2023)	Malaysian secondary schools	cross-sectional questionnaire survey	300	Metacognition/strategy awareness
S11	Durrotus et al. (2023)	Indonesian pesantren	podcast intervention report	NR	Podcast/audio
S12	Malik et al. (2023)	Arabic classroom, Indonesia	classroom action research (three cycles)	NR	Learner-centred instruction
S13	Afyuddin & Maarif (2023)	Arabic education students	learner-strategy investigation	NR	Learning strategies
S14	Fikri et al. (2024)	Indonesian university	one-group pre/post intervention	NR	Digital audio editing
S15	Ubaidillah et al. (2024)	AFL classroom, Indonesia	experimental study (publisher label)	NR	Interactive online learning
S16	Hamad & Younes (2024)	Turkish universities	descriptive questionnaire survey	127	Metacognitive / listening strategies
S17	Azka et al. (2024)	AFL learners	quantitative perception survey	NR	Barriers and access

ID	Study	Context	Design (reported)	Sample*	Primary domain
S18	Masdawati & Arifin (2024)	Indonesian senior high school	qualitative description	NR	YouTube classroom practice
S19	Tampubolon & Alee (2024)	Indonesian Islamic senior high school	classroom implementation study	NR	Eclectic instruction
S20	Habibi et al. (2024)	AFL teaching settings	applied podcast study	NR	Podcast/audio
S21	Nestia et al. (2025)	Indonesian Islamic senior high school	Quasi-experiment; non-equivalent control group	72 intervention; 36 comparison)	Interactive digital media
S22	Sabana et al. (2025)	Indonesian Islamic school	applied humanistic pedagogy study	NR	Humanistic/contextual pedagogy
S23	Hassan & Shehata (2025)	Turkish theology preparatory year (A1)	single-group pre/post task-based program	30	Task-based instruction
S24	Hikmawati et al. (2025)	Indonesian undergraduate learners	qualitative instrumental case study	39	Gamification/critical thinking
S25	Maulana et al. (2025)	Indonesian Arabic learners	media design / development	NR	Animation and gamification
S26	Nikhlattunnuha et al. (2025)	Public online Arabic media	qualitative content analysis	media corpus	Authentic online audio
S27	Othman et al. (2025)	Malaysian university	questionnaire-based evaluation	111	YouTube / e-learning
S28	Ramadona et al. (2025)	Indonesian junior secondary school	mixed-method pre/post classroom	NR	Collaborative tasks
S29	Musyarofah et al. (2025)	Indonesian university preparation program	qualitative observations descriptive	NR	Project-based instruction
S30	Rahmawati & Mustofa (2026)	Indonesian university	mixed-method questionnaire/interview	59 survey; 6 interviews	Multimodal gamification
S31	Salsabila et al. (2026a) (2026)	Indonesian university	digital learning study	NR	Digital quizzes/assessment
S32	Rosyad et al. (2026)	Indonesian intermediate AFL university	mixed-method / quasi-experimental	30	Authentic media / corpus
S33	Sofian et al. (2026)	Indonesian undergraduates	one-group pre/post over four weeks	60	AI-based microlearning

ID	Study	Context	Design (reported)	Sample*	Primary domain
S34	Mustori & Kurniawan (2026)	Indonesian higher education	qualitative instructional analysis	NR	Direct/audio-lingual instruction
S35	Rafika et al. (2026)	Indonesian secondary school	R&D with pilot trial	NR	Storytelling and animation
S36	Mubarokah & Asy'ari (2026)	Indonesian pesantren	one-group pre/post	45	Mobile app
S37	Akbar & Azhar (2026)	Indonesian primary school	one-group pre/post	26	Audio-lingual approach
S38	Salsabila et al. (2026b) (2026)	Indonesian junior secondary	qualitative classroom case	30	Gamification/Wordwall
S39	Anggraeny et al. (2026)	Indonesian senior high school	qualitative teacher interviews	NR	Digital media infrastructure
S40	Anggraini & Wahyudi (2026)	Indonesian Islamic school	reported study quasi-experimental	NR	Audiolingual practice
S41	Ilham et al. (2026)	Indonesian upper secondary	pre-experimental report	32	Direct instruction

Note. NR = sample size not confirmed from the accessible article page or abstract. *Counts are not comparable across classroom, survey and development designs.

Table 2. Abstract-level reporting and limits of inference for each provisionally charted primary study (n = 41).

Study	Publisher-visible finding / evidence object	Key interpretation boundary
S01 Rochma et al. (2021)	Describes the implementation of the Al Jazeera learning site in university listening classes.	No independent treatment effect established.
S02 Chamidah & Istiqomah (2021)	Observed stories, music, and conversation activities for listening and speaking.	Multi-skill case study; no scored listening contrast.
S03 Berutu et al. (2022)	Documents the integration of audio and images through PowerPoint for listening instruction.	Descriptive implementation, not efficacy testing.
S04 Rahmah & Ahsanuddin (2022)	The podcast product received an average feasibility judgment of 93.7% from expert and student evaluation.	Feasibility is not measured listening gain.
S05 Fawzani et al. (2022)	Reports a correlation ($r = .722$) between Arabic-song engagement and listening-skill measure.	Association cannot show song-induced learning.
S06 Kusumadewi et al. (2022)	Abstract reports different N-gain values (0.310 versus 0.242) across media conditions.	Very small sample; contradictory significance wording requires full-text checking.

Study	Publisher-visible finding / evidence object	Key interpretation boundary
S07 Mahbub & Khusnul (2022)	Reports observations and interviews on locally developed audiovisual material.	Outcome scale and comparison are unclear.
S08 Subakir et al. (2023)	Reports a 60.12% improvement in its measured listening outcome following a developed materials package.	Meaning and calculation of improvement require full-text audit.
S09 Rosyad et al. (2023) (2023)	Proposes using natural communication situations as a basis for listening material design.	Instructional design is distinct from demonstrated proficiency gain.
S10 Abdul Rahman & Baharudin (2023)	Examines metacognitive awareness and listening achievement in a sample of Malaysian students.	Association and awareness measures are not strategy-training effects.
S11 Durrotus et al. (2023)	Examines podcasts in the context of Arabic listening instruction at Al-Yasini.	Effect estimation and comparison require full text.
S12 Malik et al. (2023)	Reports repeated classroom cycles and improvement in listening scores and engagement.	Cycle changes lack controlled counterfactual.
S13 Afyuddin & Maarif (2023)	Investigates learners' Arabic listening strategies.	Sample and instruments must be verified from full text.
S14 Fikri et al. (2024)	Reports a 14.3-point mean pre/post difference after using Audacity-developed materials.	No contemporaneous control group.
S15 Ubaidillah et al. (2024)	Evaluates the alefbata.com learning platform against Arabic listening tasks.	Design, allocation, and measure require full-text confirmation.
S16 Hamad & Younes (2024)	Reports stronger use of pre- and during-listening than post-listening strategies.	Self-reported strategy use is not an objective score.
S17 Azka et al. (2024)	Finds learners report fast speech, unfamiliar vocabulary, concentration and audio-processing challenges.	Challenges are perception-reported, not independently measured difficulty.
S18 Masdawati & Arifin (2024)	Describes student activities and perceived benefits of YouTube listening practice.	Observation and interviews do not isolate medium effect.
S19 Tampubolon & Alee (2024)	Describes a combination of listening-teaching approaches in a boarding-school context.	Needs task-level assessment details.
S20 Habibi et al. (2024)	Explores podcasts as a resource for paired listening and speaking learning.	Listening-specific measure and comparison require full text.
S21 Nestia et al. (2025)	Compared Lumio-assisted listening with conventional instruction; publisher abstract reports experimental-group means increasing from 48.61 to 81.39.	Nonrandom allocation; between-group estimates, assessment reliability and baseline comparability require full-text appraisal.
S22 Sabana et al. (2025)	Analyzes a humanistic framework for active engagement in Arabic listening instruction.	Evidence of score change uncertain at abstract level.

Study	Publisher-visible finding / evidence object	Key interpretation boundary
S23 Hassan & Shehata (2025)	Reports improvement on author-developed Arabic listening subskill tests after a task-based program.	No comparison group; task/rubric validity needs checking.
S24 Hikmawati et al. (2025)	Documents Kahoot activities involving songs, dialogues, short films and stories.	Qualitative indicators of critical thinking are not treatment effect sizes.
S25 Maulana et al. (2025)	Develops multimedia listening learning using animation combined with Kahoot.	Usability and motivation should not be conflated with comprehension.
S26 Nikhlatunnuha et al. (2025)	Codes characteristics of Arabic-listening content on a podcast YouTube channel.	No learner test or language achievement outcome.
S27 Othman et al. (2025)	Surveys learners' engagement and perceived improvement for e-Kalam YouTube materials.	Reported perceived gains are not blinded listening-test effects.
S28 Ramadona et al. (2025)	Reports mean classroom scores increasing from 42.5 to 83% after collaborative listening lessons.	Uncontrolled change and assessment details require audit.
S29 Musyarofah et al. (2025)	Documents how project-based learning is implemented during Arabic-listening lessons.	Descriptive study does not yield an effectiveness contrast.
S30 Rahmawati & Mustofa (2026)	Reports positive student perceptions of YouTube-Kahoot integration in listening.	Learner satisfaction is not proof of durable comprehension.
S31 Salsabila et al. (2026a) (2026)	Examines use of Quizalize activities to support listening.	Comparator and objective score not established by accessible metadata.
S32 Rosyad et al. (2026) (2026)	Reports improved listening and word recognition with authentic news/podcast exposure.	Small, local sample and unclear allocation require audit.
S33 Sofian et al. (2026)	Reports statistically significant listening improvement after AI-supported vlog microlearning.	Its one-group design cannot isolate AI as the cause.
S34 Mustori & Kurniawan (2026)	Documents direct instruction, audiolingual exercises, and classroom drills.	Observational evidence only.
S35 Rafika et al. (2026)	Develops storytelling-based animated listening materials with expert review.	Feasibility and in-class pilot are not causal proof.
S36 Mubarakah & Asy'ari (2026)	Reports listening pre/post measurements after use of the Teach Me Arabic application.	No comparison group.
S37 Akbar & Azhar (2026)	Reports listening scores before and after audiolingual practice.	Single class; counterfactual absent.
S38 Salsabila et al.	Describes audio replay, learner responses and Wordwall task engagement.	Cannot infer treatment superiority.

Study	Publisher-visible finding / evidence object	Key interpretation boundary
(2026b) (2026)		
S39 Anggraeny et al. (2026)	Interviews teachers about digital media implementation for Arabic listening instruction.	No independent learner performance estimate.
S40 Anggraini & Wahyudi (2026)	Reports study of auditory-language method and Grade 9 Arabic listening.	Precision of design and outcome requires full-text confirmation.
S41 Ilham et al. (2026)	Reports assessment of direct-method teaching for both listening and speaking.	Publisher abstract mixes one-group design with independent-group language; verify before inference.

Note. The entries reproduce the level of inference supportable from publisher metadata/abstracts; they are not independent replications or validated effectiveness estimates. Exact outcome wording and precision require full-text audit.

4.1. Populations, designs and representativeness

Table 1 demonstrates heterogeneity at three connected levels. First, educational level varies: some studies focus on children first learning to decode short Arabic speech, while others examine university students in a language-specialist program. Second, the reported research designs range from case descriptions and media development to cross-sectional questionnaires and one-group pre/post testing. Third, the languages and registers of the recordings are not consistently visible at the abstract level. These differences impede a direct ranking of interventions.

Hassan and Shehata (2025) investigated a task-based program for 30 CEFR-A1 Turkish theology preparatory learners using a one-group pre/post design. Hamad and Younes (2024) separately surveyed 127 learners about strategies before, during and after listening and reported lower post-listening use. A Malaysian survey of 300 students examined metacognitive awareness and achievement (Abdul Rahman & Baharudin, 2023). These three examples map different evidence objects—change within a group, strategy self-report and awareness/achievement relationships—and must not be ranked on one effectiveness scale.

The mapped dataset is heavily concentrated in Indonesian institutions, partly because targeted journal-platform searching located more Indonesian records. This is a search-coverage observation only. Neither the proportion of all Arabic-listening papers published in Indonesia nor the prevalence of any instructional approach can be estimated until comprehensive multilingual database searching is complete. Figure 2 therefore depicts candidates by year, not the total field.

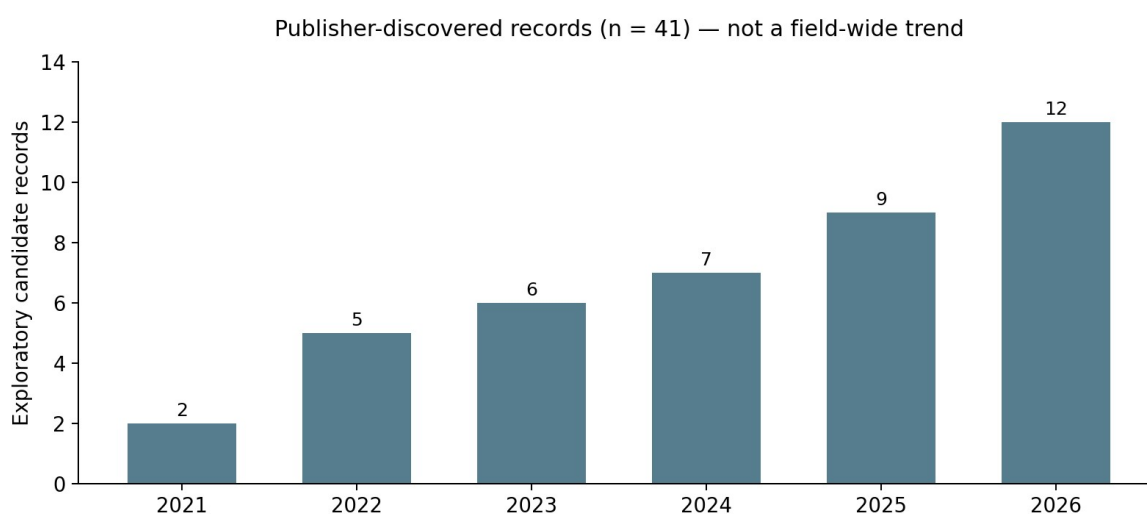


Figure 2. Year-by-year distribution of the 41 provisional publisher-discovered records. The chart must not be interpreted as a bibliometric publication trend.

The concentrated recent entries in Figure 2 are better interpreted as a screening priority than a demonstrated increase in research productivity. Final extraction should separately identify record retrieval date, actual publication year, peer-review status and overlapping samples to avoid double counting.

4.2. Listening constructs and assessment measures

A first division runs between auditory word recognition and higher-level comprehension. Songs, listening drills and short dialogues may assess sound recognition, simple word meanings or sentence responses. Authentic news, vlogs and podcasts may require sustained concentration, identification of discourse cues,

inference and tracking of the speaker's intent. It is unsound to compare these instruments only because all are called tests of *mahārah al-istimāʿ*.

Published abstracts seldom provide the exact listening test, audio transcript, item difficulty analysis, internal consistency, piloting procedure or a statement about the use of Modern Standard Arabic versus dialect. The 2022 study of ten elementary learners by Kusumadewi et al. (2022), for example, reports modest and differing N-gain values for audiovisual comparisons, but the abstract contains wording that merits checking against the full paper. A study reporting expert feasibility judgments of 93.7% for Spotify podcast materials is answering a different question from a study measuring achievement after listening exposure (Rahmah & Ahsanuddin, 2022).

Table 2 distinguishes media feasibility, user perceptions, observational descriptions, associations, pre/post change and controlled comparisons. For development research, material clarity, usability and comprehension need separate extraction fields; for surveys, beliefs require separation from independent performance; and for interventions, the original scales, baseline scores, variance, test equivalence, attrition, treatment dose and delayed measures require confirmation. Figure 5 operationalizes these construct-to-outcome distinctions, which should be cross-checked against individual full texts.

4.3. Auditory and audiovisual input

The digital media subset extends from institutionally curated content to open, authentic content. Rochma et al. (2021) explored Learning Aljazeera.Net at a university; Berutu et al. (2022) described audio-visual integration through presentation software; and Fikri et al. (2024) tested materials edited using Audacity. Audio-editing tools can improve signal-to-noise quality, but editing, scaffolding and exposure time must be separated if the goal is to attribute learning changes.

A recurring input-design question concerns authenticity. Podcasts and news can increase naturalistic lexical density and pragmatic context, but recordings that are too fast, colloquial or acoustically complex may overload beginners. Azka et al. (2024) document learner-reported problems associated with fast speech, unfamiliar vocabulary and attention. More recent corpus-informed work reports examining lexical density and discourse markers in Arabic news/podcasts with classroom tests (Rosyad et al., 2026). The latter is a promising bridge between corpus description and pedagogy, but its allocation and measurement details require full-text confirmation.

Audiovisual media complicate the outcome even further because learners may answer by reading subtitles or interpreting pictures rather than understanding the audio stream. The appropriate transfer test for a listening claim would include a novel, audio-only passage with the same target listening construct. The gain from multimodal comprehension under supported conditions is important, but should be named as supported comprehension rather than independent auditory ability.

4.4. Task-based, collaborative and learner-centred approaches

Task-based teaching seeks to link listening to purposeful comprehension and communication. Hassan and Shehata (2025) built a program around functional listening needs for 30 Turkish preparatory-year students and reported improved outcomes within a one-group design. By contrast, Ramadona et al. (2025) examined

collaborative use of dialogic material in an Indonesian junior secondary school, with the paper reporting mean scores of 42.5 and 83 across pre/post assessments. Neither result alone establishes a superiority claim against alternatives without appropriate comparison conditions and transparent scoring.

Other descriptive research foregrounds learner agency: activities involving stories and conversation (Chamidah & Istiqomah, 2021); reflective humanistic participation (Sabana et al., 2025); project-based listening activities (Musyarofah et al., 2025); and combined direct, audiolingual and eclectic routines (Tampubolon & Alee, 2024; Mustori & Kurniawan, 2026). The explanatory objective of these studies is often local implementation rather than a controlled treatment test. The appropriate synthesis preserves that distinctive contribution.

A well-specified intervention report would reveal audio text complexity, item types, the number of listening passes, whether students could ask for clarification, group size, teacher mediation and the relationship between the intervention task and the posttest. If the same audio is used repeatedly for training and testing, observed improvement may reflect passage familiarity rather than generalizable listening development.

4.5. Metacognition, anxiety, strategy use and feedback

Strategy research raises questions that technology evaluations frequently leave implicit: do learners predict content from context, monitor their comprehension, allocate attention to key information and review errors after listening? Hamad and Younes (2024) found that the survey respondents reported stronger pre- and during-listening practice than post-listening practice. This finding does not estimate which strategy raises proficiency, but it identifies an actionable area for instruction and subsequent experimental testing.

The relationship between listening and songs is another example of how an attractive pedagogical medium can be supported by several types of evidence. Fawzani et al. (2022) reported $r = .722$ between Arabic songs and a listening-related measure among 68 university students. That correlation cannot demonstrate that songs caused development. Future studies might randomize exposure dosage, equate listening tests and include affective ratings and delayed novel-content tests to disentangle motivation from learning.

Perceptual difficulties should be defined at task level. A learner who struggles with rapid speech may lack lexical knowledge, automatic segmenting skill, attention control or experience with the chosen dialect. Survey summaries describing obstacles (Azka et al., 2024) are useful hypothesis generators but are not substitutes for acoustic manipulations or validated diagnostic instruments. Figure 3 supports identifying these hypothesized mechanisms and matching them to measures.

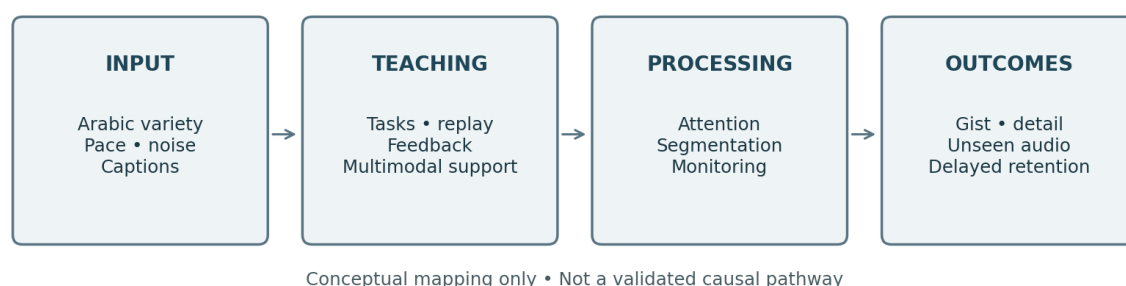


Figure 3. Proposed construct-alignment framework linking Arabic listening input, pedagogical mediation, learner processing and assessment; causal paths are hypotheses, not established review findings.

Figure 3 is an interpretive framework rather than a causal diagram supported by pooled data. The final review should chart each component rather than collapse them under a single 'listening score'. For example, an improvement in vocabulary recognition cannot be presumed to represent enhanced inferential comprehension, and a multimedia-assisted response may not transfer to unfamiliar audio-only conditions.

4.6. Gamification, platform integration and AI-mediated listening

Recent platform-based work includes Kahoot-supported comprehension tasks (Hikmawati et al., 2025), animated media with Kahoot (Maulana et al., 2025), YouTube/Kahoot integration (Rahmawati & Mustofa, 2026) and applications such as Quizalize and Wordwall (Salsabila, S., et al., 2026; Salsabila, N. A., et al., 2026). Platforms should be coded by their learning function—delivery, replay, feedback, response collection or motivation—not treated as intrinsically comparable interventions. Figure 4 differentiates the kinds of evidence such studies supply.

Hikmawati et al. (2025) studied 39 undergraduates in a qualitative single-case design, documenting critical-thinking activity rather than a controlled listening effect. Othman et al. (2025) surveyed 111 learners about a YouTube learning kit, with perceptual outcomes. In a stronger comparative design than suggested by the original draft, Nestia et al. (2025) used a non-equivalent-control-group quasi-experiment: 36 students received Lumio-supported instruction and 36 formed a comparison group. The article abstract reports experimental-group means changing from 48.61 to 81.39, but baseline equivalence, exact between-group effects and measurement validity should still be extracted from its full text.

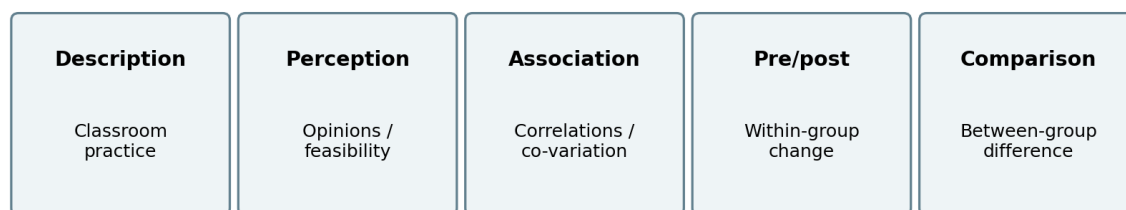


Figure 4. Evidence-type taxonomy: classroom description, perceptions and feasibility, association, pre/post change, and controlled contrast are not interchangeable.

Figure 4 contextualizes the positive findings reported by platform studies. A percentage derived from product feasibility, for example, cannot be interpreted as an equivalent percentage improvement in independent listening comprehension.

AI-adaptive systems introduce additional reporting demands. Sofian et al. (2026) studied a vlog-based AI microlearning program with 60 undergraduates over four weeks in a one-group pre/post design. Positive change in its listening test is an important report-level outcome but does not identify the independent causal contribution of the AI element, the short-video modality or additional practice. Future studies should document tool/model version, prompts, personalization logic, user data safeguards, audio generation provenance and whether students improve on independent listening materials without tool support.

5. Discussion

5.1. Central synthesis: construct alignment before method ranking

The principal contribution suggested by this preliminary map is the need for construct-aligned interpretation. Research labelled 'Arabic listening' spans a continuum from hearing and discriminating phonological forms to comprehending sustained, authentic speech. A basic audio discrimination task can be a valid achievement measure in a beginner curriculum but cannot be used to claim gains in academic lecture listening. Conversely, a survey of strategy use can illuminate self-regulation without measuring auditory decoding. The draft therefore refrains from ranking YouTube, podcasts, audiolingual instruction, task-based learning or AI as definitively 'best' for Arabic listening.

The framework in Figure 3 allows a more coherent comparison. The input layer specifies what was heard; the pedagogy layer identifies what learners were asked to do; the processing layer proposes cognitive engagement; and the outcome layer identifies what evidence was collected. This specification also makes context visible: the same news podcast may provide comprehensible material to advanced learners but overwhelm a beginning class. Age, native language, prior script literacy and variety of Arabic may influence interpretation across all layers.

The candidate findings in Table 2 should consequently be read as a mapped range of outcomes rather than a vote count of positive studies. Publishers often report improvement without enough abstract detail about allocation, instrument validation or test reuse. A scoping review can highlight that pattern as a reporting gap without converting every reported gain into causal evidence.

5.2. Validity threats in the available record

Three threats to inference must be kept distinct. One-group pre/post reports permit maturation, test familiarity, changes in exposure and instructor effects as alternatives to the proposed intervention effect (Hassan & Shehata, 2025; Sofian et al., 2026). Media feasibility and survey ratings measure acceptability or reported experience rather than auditory performance. Controlled studies, including the non-equivalent-group Lumio report (Nestia et al., 2025), offer an explicit comparator but still require scrutiny of group comparability, selection and outcome

calibration. Figure 4 positions these report types as different evidence objects, not interchangeable treatment estimates.

Quantitative indicators should therefore be interpreted carefully. The 93.7% reported by Rahmah and Ahsanuddin (2022) describes evaluation of the developed medium, not percent increase in listening proficiency. The 14.3-point pre/post difference reported by Fikri et al. (2024) has meaning only relative to the test scale and design; it is not standardized across contexts. The association reported by Fawzani et al. (2022) is a correlation, not an intervention effect. Each numerical statement belongs to its originating assessment system.

At the meta-level, the charting process has its own weaknesses. Searches limited to a few accessible journal platforms introduce source-selection bias; a high proportion of recent Indonesian records reflects discoverability, not necessarily field geography. Title/abstract metadata may contain inconsistencies in authorship, statistics and date metadata. A defensible final review must retrieve complete articles, audit numerical values and preserve reasons for noninclusion.

5.3. Priority research and reporting agenda

Future research should specify audio language variety, native-speaker status, speech rate, recording conditions, length, genre, subtitle availability, number of replays and task support. Listening instruments should separately measure phonetic/lexical recognition, gist/detail, inference, and transfer to novel audio-only stimuli; Figure 5 proposes a practical charting structure. Where possible, study authors should report test blueprints, parallel forms, rater/scorer checks and delayed retention.

Construct	Outcome distinction	Suggested measurement
Phonetic / word level	Sound discrimination · lexical segmentation	Novel words in audio-only clips
Sentence / passage level	Gist · details · inferential comprehension	Parallel unseen passages
Independent transfer	Unscaffolded listening · delayed retention	Novel genre after a delay

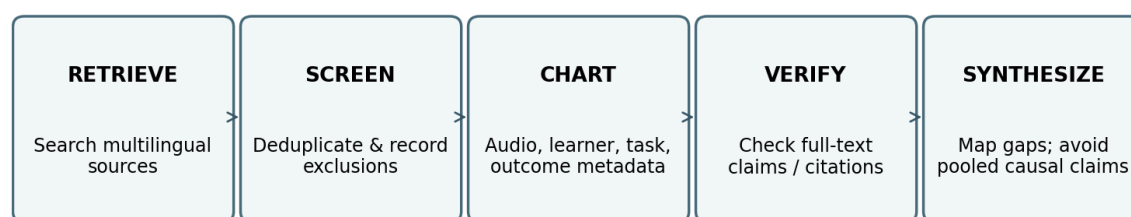
Figure 5. A construct-to-measurement matrix for lexical, passage-level and independent-transfer outcomes in Arabic listening research.

As Figure 5 indicates, novel passage-level and delayed audio-only tasks provide different evidence from immediate word-identification exercises; the instrument must reflect the outcome claimed.

Research designs should be selected to match the intended claim. Ethnography, case study and classroom-action designs are well suited to studying local practices and learner experiences. They should not be assessed exclusively as failed experiments. However, studies claiming comparative instructional effectiveness require credible allocation, baselines, suitable comparison conditions, timing, assessment blinding and attrition reporting. For digital media, separate the pedagogical principle from

the brand or platform, and specify what the technology contributes beyond increased exposure.

A reproducible multilingual search is essential. Database indexing and Arabic-language records, Turkish and Malaysian research, Middle Eastern and North African contexts, heritage-learning boundaries and uneven access to technology could substantially change the map. Journal prestige should not be inferred from DOI assignment. Figure 6 sets out the outstanding search–screen–chart–verify–synthesize sequence; each step is auditable and must be completed before final PRISMA-ScR reporting.



A reproducible process is a precondition for calling the review completed

Figure 6. Proposed completion workflow for a reproducible scoping review; the steps have not yet been completed for this draft.

The sequence in Figure 6 defines a submission threshold rather than an assertion that missing database and reviewer records were reconstructed.

5.4. Limitations of this initial manuscript

This manuscript remains an exploratory evidence map. No subscription-database export files, dual screening log, complete exclusion reasons, or fully reconciled record-level outcome extractions were supplied. Consequently the 41 candidates must not be described as final eligible studies and the count in Figure 2 cannot support bibliometric inference. Multiple bibliographic entries and quantitative claims still require full-text checking, notwithstanding the publisher-level confirmation available for selected studies.

The temporal distribution is selection-dependent. Study locations and sample populations may be incompletely described or concealed by English abstracts. Although both tables contain one row per record, eligibility decisions can alter their contents. Accordingly, Figure 1 contains no fictional counts, Figure 2 cannot be used as a publication-trend estimate, and no pooled treatment conclusion is supported. Any submission claiming completed PRISMA-ScR compliance must await reproducible search execution and additional review work.

6. Conclusion

The exploratory literature spans task-based Arabic listening, authentic audio, multimedia resources, songs, podcasts, gamified practice, metacognitive strategies and AI-supported microlearning. The contribution of a completed scoping review would be to demonstrate which listening constructs are targeted, how input and learner conditions shape assessment, and where credible comparisons or transfer measures are missing—not to name a universally superior platform. An auditable

multilingual search, two-reviewer screening, full-text verification and finalized PRISMA-ScR diagram remain prerequisites for claiming a completed review. Figure 6 specifies the steps needed to achieve that standard.

Declarations

Protocol registration: None reported. Funding: To be confirmed by the authors. Conflicts of interest: Author declarations pending. Ethics: No new participants were recruited for this manuscript revision. Data availability: The cited DOI links support record location; raw database exports, screening decisions and full-text extraction sheets remain unavailable. Authorship, affiliations and acknowledgments must be completed by the authors.

References

The references below retain the draft's 47-source working corpus. DOI links facilitate verification but do not establish that each source has been audited at full-text level. Publisher-record corrections are integrated where independently supported; unresolved authorship, pagination and outcome-claim discrepancies are itemized in the separate reference-validation document.

- Abdul Rahman, N. F., & Baharudin, H. (2023). Metacognitive awareness level and achievement in Arabic listening skills among secondary school students. *International Journal of Academic Research in Business and Social Sciences*, 13(12). <https://doi.org/10.6007/IJARBS/v13-i12/20197>
- Afyuddin, M. S., & Maarif, M. S. (2023). Listening learning strategy used by Arabic education students. *Al-Irfan*, 6(1), 138–166. <https://doi.org/10.58223/alirfan.v6i1.6617>
- Akbar, M., & Azhar, M. (2026). Efektivitas metode audio lingual dalam meningkatkan maharah istima' siswa kelas IV SDIT Imam Ahmad. *Pendas*, 11(3). <https://doi.org/10.23969/jp.v11i03.66656>
- Anggraeny, R., Yunita, N., Usshofia, A., & Sabilla, I. R. (2026). Pengembangan media pembelajaran bahasa Arab berbasis digital untuk meningkatkan keterampilan maharah istima' di MA Pembangunan UIN Jakarta. *Pendas*, 11(3). <https://doi.org/10.23969/jp.v11i03.63343>
- Anggraini, E. A., & Wahyudi, M. (2026). أثر تطبيق المنهج السمعي اللغوي في مهارة الاستماع في اللغة العربية لدى طلاب الفصل 9.1 بمعهد الاتفاقية. *Lughoti*, 9(2). <https://doi.org/10.53649/lughoti.v9i02.2287>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Azka, L., Hafsoh, U. N., & Ibrahim, F. M. A. (2024). The challenges of learning Arabic listening skills using YouTube as a learning tool. *Alsina: Journal of Arabic Studies*, 6(2). <https://doi.org/10.21580/alsina.6.2.23369>
- Berutu, W. R., Shodiq, M. J., & Yuniar (2022). Utilization of Microsoft Power Point media in Mahārāh Istimā' (Arabic learning skill) learning in the digital age. *EduLab*, 7(1), 21–36. <https://doi.org/10.14421/edulab.2022.71.02>
- Chamidah, D., & Istiqomah, R. (2021). Linguistic activities for improve listening and speaking skills. *Ijaz Arabi Journal of Arabic Learning*, 4(2), 209–227. <https://doi.org/10.18860/ijazarabi.v4i2.10528>
- Durrotus, Rahmatillah, R., Alifah, & Taufiqurrochman, R. (2023). Pengaruh media podcast dalam meningkatkan maharah istima' santri Pondok Pesantren Al-

- Yasini. Shaut al Arabiyyah, 11(1), 104–111. <https://doi.org/10.24252/saa.v11i1.25597>
- Fawzani, N., Firdaus, & Akmal (2022). Hubungan lagu berbahasa Arab dengan maharah istima' mahasiswa. Jurnal Naskhi, 4(2), 32–39. <https://doi.org/10.47435/naskhi.v4i2.1214>
- Fikri, S., Kirom, M. U., & Al-Furqaan, D. L. (2024). Audacity: Learning media design in listening skills for Arabic Special Development Program. International Journal of Arabic Language Teaching, 6(1), 1–18. <https://doi.org/10.32332/ijalt.v6i01.7935>
- Habibi, M. T., Hanifansyah, N., Solehudin, M., Mahmudah, M., & Syakur, S. A. (2024). Podcasts as an innovative solution for teaching Arabic: Enhancing speaking and listening skills. Studi Arab, 15(2), 87–105. <https://doi.org/10.35891/sa.v15i2.5784>
- Hamad, M. M., & Younes, H. (2024). Learning strategies for listening skills among Arabic learners in the preparatory year at faculties of Islamic sciences in Turkey. Ijaz Arabi Journal of Arabic Learning, 7(3). <https://doi.org/10.18860/ijazarabi.v7i3.29340>
- Hassan, A. M., & Shehata, H. (2025). The effectiveness of a task-based program in developing Arabic listening skills for non-native speakers. Ijaz Arabi Journal of Arabic Learning, 8(3). <https://doi.org/10.18860/ijazarabi.v8i3.34542>
- Hikmawati, S. A., Fatim, A. L. N., Hamid, M. A., Ma'arif, A. S., & Bakhiet, B. M. (2025). Kahoot gamification: Fostering critical thinking in Arabic listening comprehension. Alsina: Journal of Arabic Studies, 7(1), 29–62. <https://doi.org/10.21580/alsina.7.1.28086>
- Ilham, Haniah, & Chalikh, S. A. (2026). Efektivitas metode mubasyarah dalam meningkatkan maharah istima' dan maharah kalam pada pembelajaran bahasa Arab di PDF Ulya' DDI Mangkoso. Journal of Arabic Education and Linguistics, 5(2). <https://doi.org/10.24252/jael.v5i2.55716>
- Kusumadewi, I. P., Alwi, I. M., & Yekti, S. (2022). Efektivitas media audio visual terhadap keterampilan istima' di MI Muhammadiyah Gembuk 1. Jurnal Naskhi, 4(2), 1–9. <https://doi.org/10.47435/naskhi.v4i2.1222>
- Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. Implementation Science, 5, 69. <https://doi.org/10.1186/1748-5908-5-69>
- Mahbub, M., & Khusnul, R. (2022). Pengembangan media audio visual untuk pembelajaran maharah istima' di Madrasah Tsanawiyah Al-Amiriyah Blokagung Banyuwangi. Tadris Al-Arabiyyah, 2(2). <https://doi.org/10.30739/arabiyah.v2i2.1629>
- Malik, I., Hairuddin, Miolo, M. I., & Samad, M. A. (2023). Silent way method sebagai alternatif pembelajaran maharatul istima'. Al-Kilmah, 2(2). <https://doi.org/10.58194/alkilmah.v2i2.1850>
- Masdawati, L. J., & Arifin, M. B. U. B. (2024). Utilization of YouTube media in training Maharah Istima' at Muhammadiyah Sidoarjo High School. Frasa, 5(2), 37–47. <https://doi.org/10.36232/frasaunimuda.v5i2.900>
- Maulana, M. R., Darmawaty, R., & Anaam, N. T. F. (2025). Joyful listening: Integrating 'Nusa dan Rara' animation and KAHOOT! for maharah istima'. Jurnal Al Bayan, 17(1), 159–177. <https://doi.org/10.24042/9rf63d07>
- Mubarokah, A., & Asy'ari, H. (2026). Pengaruh aplikasi Teach Me Arabic terhadap peningkatan maharah istima' santri. Pendas, 11(3). <https://doi.org/10.23969/jp.v11i03.61115>

- 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, 143. <https://doi.org/10.1186/s12874-018-0611-x>
- Mustori, & Kurniawan, W. (2026). Arabic language learning strategies in enhancing students' maharah istima' at Al-Ghurabaa Islamic Institute Jakarta. *Fikroh*, 19(2). <https://doi.org/10.37812/fikroh.v19i2.2373>
- Musyarofah, B., Halifah, N., Sisran, & Fitri, Y. N. (2025). Project based learning (PBL) to improve listening skills in Arabic language learning. *Indonesian Journal of Education and Social Studies*, 4(3). <https://doi.org/10.33650/ijess.v4i3.13159>
- Nestia, B., Bukhori, E. M., & Huda, M. M. (2025). Integration of interactive Lumio media for the development of Arabic listening skills in Islamic senior high schools. *Arabiyatuna: Jurnal Bahasa Arab*, 9(1), 1–26. <https://doi.org/10.29240/jba.v9i1.12441>
- Nikhlatunnuha, D., Zaenuri, M., Islam, S., & Amar, M. (2025). Digital education on YouTube: An analytical study of Arabic language learning content on the 'Arab Podcasts' channel on YouTube. *Al-Muyassar*, 4(2). <https://doi.org/10.31000/al-muyassar.v4i2.13521>
- Othman, S., Arshad, A., Fauzi, A., & Mohd Zaidin, N. A. (2025). Exploring the integration of YouTube-based learning to improve Arabic learners' listening and speaking skills. *International Journal of Research and Innovation in Social Science*, 9(9), 4926–4936. <https://doi.org/10.47772/IJRISS.2025.909000401>
- Rafika, Imron, K., & Nazarmanto (2026). Pengembangan media pembelajaran maharah istima' berbasis storytelling di MTs Zaadul Maad Palembang. *Pendas*, 11(2). <https://doi.org/10.23969/jp.v11i2.47743>
- Rahmah, L. A., & Ahsanuddin, M. (2022). Pengembangan media podcast pada aplikasi Spotify sebagai media pembelajaran maharah al-Istima'. *Journal of Language Literature and Arts*, 2(11), 1613–1625. <https://doi.org/10.17977/um064v2i112022p1613-1625>
- Rahmawati, B. L., & Mustofa, S. (2026). Persepsi mahasiswa terhadap penggunaan integrasi Youtube dan Kahoot di jurusan Pendidikan Bahasa Arab Universitas Islam Negeri Maulana Malik Ibrahim Malang. *Al-Wasil*, 4(1). <https://doi.org/10.30762/alwasil.v4i1.8709>
- Ramadona, Mukmin, & Sabana, R. (2025). Pembelajaran maharah istima' menggunakan kitab Muhawaroh Haditsah dengan pendekatan collaborative learning di MTs An-Nuur Palembang. *Pendas*, 10(4). <https://doi.org/10.23969/jp.v10i04.38285>
- Rochma, S. N., Mahmudah, U., & Yahya, Y. K. (2021). Utilizing technology in Arabic teaching: Implementation of media Learning Aljazeera.Net on listening skill teaching at University of Darussalam Gontor. *Arabiyatuna: Jurnal Bahasa Arab*, 5(2). <https://doi.org/10.29240/jba.v5i2.2730>
- Rosyad, M. S., Ibrahim, F. M. A., & Bahrudin, U. (2023). The natural situation and its effective role in designing Arabic listening skills teaching materials for non-native speakers. *Ijaz Arabi Journal of Arabic Learning*, 6(3). <https://doi.org/10.18860/ijazarabi.v6i3.19531>
- Rosyad, M. S., Nuruddin, A., Makinuddin, M., Rozaq, M. A., & Hafidz, M. (2026). Digital authentic media exposure and listening development in Arabic as a foreign language: A corpus-informed approach. *Borneo Journal of Language and Education*, 6(2). <https://doi.org/10.21093/benjole.v6i2.12811>
- Sabana, R., Zaini, H., Nurcholis, A., Daniel, S. M., & Hidayah, N. (2025). Learning istima' based on Ibnu Tufail's humanism approach at Applied Islamic School

- Prof. Muhajrin Palembang. Arabiyatuna: Jurnal Bahasa Arab, 9(1), 249–266. <https://doi.org/10.29240/jba.v9i1.12027>
- Salsabila, S., Rofiqi, M., Sriyono, S. A., Shodiq, M. J., & Muliati, A. Y. (2026). The use of Quizalize media to improve Arabic listening skills of students at Nahdlatul Ulama University. Mantiqu Tayr, 6(1). <https://doi.org/10.25217/mantiqu tayr.v6i1.7160>
- Salsabila, N. A., Hanifah, U., & Astutik, N. S. P. (2026). Pemanfaatan media pembelajaran interaktif berbasis Wordwall untuk meningkatkan maharah istima' murid kelas VII-H MTsN 4 Sidoarjo. Pendas, 11(3). <https://doi.org/10.23969/jp.v11i03.66039>
- Sofian, G. Y., Sutaman, Taufiqurrochman, R., Hidayat, A. F. S., & Nukman (2026). The effectiveness of AI based microlearning platforms through vlog content to improve Arabic listening skills. Arabiyatuna: Jurnal Bahasa Arab, 10(1), 39–62. <https://doi.org/10.29240/jba.v10i1.16265>
- Subakir, F., Mufliah, L., Halim, F., Fauziah, S. A. B. N., & Adam, S. B. (2023). The effectiveness of Arabic learning materials to improve students' listening skills at Darussalam Islamic Boarding School. Jurnal Al Bayan, 15(1), 173–191. <https://doi.org/10.24042/albayan.v15i1.14442>
- Tampubolon, M. S., & Alee, A. S. (2024). Implementation of maharah istima' learning using the eclectic method in class 1 of Madrasah Aliyah Islamic Center Bin Baz Islamic Boarding School. At Turots, 6(2), 663–672. <https://doi.org/10.51468/jpi.v6i2.543>
- 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>
- Ubaidillah, U., Millah, F. I., & Sapitri, N. (2024). The use of online media 'alefbata.com' in improving Arabic listening skills: Experimental study. Al-Ta'rib, 12(1), 103–114. <https://doi.org/10.23971/altarib.v12i1.7852>
- Wahyuni, S., Ritonga, M., & Afrianti, W. (2023). Systematic review of learning method for teaching Arabic listening and speaking skills. Al-Hayat: Journal of Islamic Education, 7(1), 30–41. <https://doi.org/10.35723/ajie.v7i1.321>
- Yonan, Y., Kosim, N., & Ardiansyah, A. A. (2026). A systematic literature review of digital media in enhancing Arabic listening comprehension. Kalamuna: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban, 7(2). <https://doi.org/10.52593/klm.07.2.10>