



Digital Technologies in Qur'anic Learning: A Systematic Review of Recitation, Memorization, Educational Outcomes, and Implementation

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

Universitas Negeri Malang, Indonesia

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

Abstract

Digital tools are used to support Qur'anic reading, recitation, tajwid, and memorization, yet reported effectiveness often conflates application usability with educational improvement. This study develops a reproducible systematic-review protocol and maps 23 candidate primary articles located through targeted publisher searches. The provisional literature covers classroom media, mobile memorization, automated speech feedback, and technology adoption. Some uncontrolled before-after evaluations report short-term score increases; computational papers primarily test recognition accuracy or word error rates. Surveys and qualitative studies describe motivation, teacher readiness, infrastructure, and barriers to implementation. The evidence does not support pooling these dissimilar outcomes or claiming durable learning benefits. Comprehensive database exports, independent full-text screening, and design-sensitive risk-of-bias appraisals remain incomplete. The proposed synthesis therefore separates learner outcomes from algorithm validation and user experience, and prioritizes expert oversight, transparent evaluation, longer follow-up, and equitable access.

Keywords: mobile learning; tajwid feedback; memorization retention; technology adoption

1. Introduction

Digital media have become routine elements of educational practice, and Qur'anic learning is increasingly shaped by mobile applications, online instruction, voice recognition, and personalized feedback. The Holy Qur'an occupies a distinctive place in Muslim religious education: learning can involve decoding Arabic script, mastering the rules of tajwid, memorizing verses (hifz or tahfiz), understanding meanings, and cultivating practices of interpretation and devotion. These activities involve different learning objectives. An application that correctly plays a verse, for example, may facilitate repetition without demonstrating a durable improvement in the learner's memorization, and a speech recognizer that matches a transcript may still fail to diagnose a relevant phonetic error. An academically useful review must therefore define which educational activity and which outcome each technology is designed to support.

Digital innovation is commonly framed as a response to limited access to qualified teachers, transport constraints, uneven availability of recitation practice, and disruption during the COVID-19 pandemic. Mobile and web-based resources can provide audio models, visual cues, asynchronous activities, and opportunities to revisit material outside class. However, they also place interpretive and evaluative decisions inside devices and software systems. The quality of feedback, the reliability of Arabic script rendering, and the decision to disclose uncertainty about

pronunciation all become part of instruction. Early studies of remote memorization and online learning show why continuity and accessibility matter, but they cannot always distinguish the influence of software from concurrent changes in parental support, time available, or teacher contact (Hasjanah et al., 2022).

Research on application design and learning outcomes is growing, yet uses heterogeneous measures. A recent Android application study reported a mean increase from 65 to 85 on a locally administered memorization assessment, suggesting scope for improved practice (Djainudin et al., 2025). A different study reported an association between an iOS memorization application and learner motivation using a version of the Motivated Strategies for Learning Questionnaire, while conceding that most motivational variation lay outside the application (Hamidah & Yulius, 2024). These are not interchangeable claims: one refers to performance in an assessment, another to self-reported motivation. Neither alone establishes persistence of memorization months after the intervention, performance across all recitation contexts, or generalizability to learners with distinct Arabic fluency.

The same distinction is important in reading instruction. A development study of the Bestin application combined expert assessment and a small pupil trial; its conclusions about feasibility are useful but should be separated from evidence of causal effectiveness (Aeni et al., 2026). Classroom studies using digital projection or YouTube material similarly describe practical integration, resource availability, and student responses, yet often provide no independent control condition (Sutarmi et al., 2025). Digital delivery is not an educational mechanism by itself. At least one intervention needs to specify whether progress is expected to arise from more frequent practice, better formative feedback, enhanced motivation, greater clarity of examples, or additional contact with a teacher.

Artificial intelligence expands the distinction between technical and pedagogical performance. Recitation recognizers operate on audio signals and are assessed with accuracy, error rates, or classification metrics. Work using hidden Markov models and duration-sensitive phoneme classification shows why the pronunciation rules of Qur'anic Arabic present challenges that general speech transcription does not capture (Alqadasi et al., 2023). Later neural-recognition work compared Whisper with DeepSpeech and described word-error-rate gains on a dedicated male-and-female recitation dataset, while explicitly excluding full tajwid grading from the tested task (Alshboul et al., 2025). Neither result should be represented as evidence that children learned more effectively when using these systems. A bridge from model output to educational value requires testing the quality of corrective feedback and the response of real learners.

Conversely, the presence of advanced digital functionality does not establish widespread adoption. A survey of 205 foundation-level students reported that more than half had never used relevant artificial-intelligence applications and fewer than one in ten used them daily (Ab Alim et al., 2025). A mixed-method study of teachers' readiness for gamified AI-assisted Qur'anic learning identified comparatively strong technical readiness but weaker pedagogical readiness and limited support services (Nugraha et al., 2025). These findings are not a contradiction. They expose the gap between technology availability and the sustained practices needed to make a resource educationally meaningful.

The review must also attend to potentially adverse outcomes. Mobile devices can facilitate spaced practice and repeated listening while opening access to unrelated notifications and social-media activity. In one mixed-method study, investigators did not find a statistically significant association between reported social-media time and memorization levels, even though interview narratives raised concerns about distractions and concentration (Abdelazeez et al., 2024). A balanced synthesis needs to preserve the null quantitative finding rather than selecting only the more memorable qualitative concern. More broadly, digital inequity, privacy, the credibility of automatically generated feedback, and the replacement of individualized teacher guidance should be examined rather than dismissed as implementation details.

Source-level verification identifies three bibliographic issues requiring careful reporting: the POSITIF Android-application article was published online in February 2023 within a volume labelled 2022; the Hakimi et al. journal page lists volume 2, issue 9, while its DOI encodes volume 2, issue 8; and the Busro and Hakim article page and its PDF display different DOI suffixes. These discrepancies should be reported as metadata conflicts rather than resolved by inference.

Existing publications span several research traditions. In educational development research, investigators produce artefacts and assess usability or expert judgments; in intervention research, they compare pre- and post-performance, sometimes with a control group; in information systems, they examine use or satisfaction; in machine learning, they evaluate prediction systems with held-out audio; and in qualitative education, they study institutional constraints, perceptions, and classroom practice. These strands can answer complementary questions, but only after their evidence functions are distinguished. Pooling app acceptability with recitation accuracy or combining word-error rates with memorization scores would create an uninterpretable summary. Even within learner-outcome studies, differences in the level of teaching, frequency of exposure, skills assessed, and religious educational setting can preclude straightforward comparison.

The research problem addressed by this review is therefore not simply whether technology is “good” for studying the Qur’an. It is whether contemporary primary studies provide credible, replicable and contextually interpretable evidence for particular learning functions. The central review question is: How are digital and AI-assisted tools used in Qur’anic recitation, reading and memorization, what learner and implementation outcomes are reported, and how robust are the associated evaluation designs? Four subsidiary questions examine (RQ1) reading and recitation learning outcomes; (RQ2) memorization support and retention; (RQ3) validity of automatic pronunciation and speech feedback; and (RQ4) adoption, usability, learner engagement, equity, and pedagogical oversight. The proposed contribution is an evidence hierarchy that keeps artifact functionality, user experience, learner improvement, and generalizable effectiveness analytically separate.

This article provides the design for a focused systematic literature review and a transparent mapping of an initial group of research articles. It does not describe an exhaustive database search or claim to have established the final included-study set. Its Methods section specifies a reproducible route to PRISMA-compliant evidence identification; the preliminary findings are then organized by four substantively distinct outcome domains. The resulting Discussion specifies where research

designs require strengthening before policy, curriculum, or technology-development decisions can rely on them. Ultimately, the purpose is to support scholarly interpretation of digital Qur'anic learning without confusing technological novelty with pedagogical success.

2. Methods

2.1. Review design and review question

The review is designed as a systematic review of original empirical studies, with structured narrative synthesis planned in advance. PRISMA 2020 will guide full reporting, PRISMA-S will guide search reproducibility, and SWiM principles will support transparent synthesis when effect-size pooling is inappropriate (Page et al., 2021) (Rethlefsen et al., 2021) (Campbell et al., 2020). The population includes learners and teachers in organized or informal educational settings where the focal activity is Qur'anic reading, recitation, tajwid, memorization or meaning-related learning. The intervention or exposure is a digital application, digital learning platform, audio-feedback tool, or AI-assisted learning workflow. Comparator types may include conventional teacher-led practice, a different digital tool, a pre-intervention baseline, or no comparator in implementation studies. Outcome definitions are prespecified before final screening.

2.2. Sources and search strategy

The planned searches cover Scopus, Web of Science Core Collection, ERIC, IEEE Xplore, ACM Digital Library, and supplementary publisher or journal websites, with backward and forward citation searching. The planned time window is 1 January 2022 through 11 October 2026. Searches will use three concept blocks: Qur'anic terminology (Quran, Qur'an, Qur'anic, Al-Quran); learning processes (recitation, tajwid, tilawah, memorization, tahfiz, hifz, reading, teaching); and technology (mobile, application, digital, online, deep learning, speech recognition, AI, gamification). The precise syntax for major platforms is in Appendix C. Database-specific dates, record numbers, exported metadata and revisions are not yet available. Searches conducted for this draft were targeted discoveries of individual public journal records, not comparable, independently reproducible database-wide exports.

The final search must preserve exact query strings, date-stamped export files, database fields, records removed before screening, and full-text exclusions. Search yields were not reconstructed from the number of currently known studies, and Figure 1 must not be replaced with numerical boxes until the screening ledger is available.

2.3. Inclusion and exclusion criteria

Original journal studies published in the eligible period may be included when they examine both Qur'anic learning and a specific digital tool or digital instructional environment, and report empirical evaluation through learner assessment, survey, interviews, observations, computational benchmarking, or usability tests. Eligible full texts may be in English or Indonesian, with language recorded rather than presumed. Technical papers are eligible only when they demonstrate a plausible interface to Qur'anic recitation or memorization learning; generic Arabic speech studies without Qur'anic learning relevance are excluded. Editorials, books, preprints without a peer-reviewed counterpart, narrative reviews, systematic

reviews, conference abstracts and purely normative discussions are excluded from the included primary-study population. Studies exclusively of non-digital teaching methods are background comparators, not primary included studies.

2.4. Screening, extraction and appraisal

English and Indonesian full texts are eligible under the current language criterion; the 2025 QURANICA record identifies its article type as Malay, although an English abstract is available. The full paper will therefore need a documented eligibility decision or a preregistered change to the language rule before entering the final included-study corpus.

Two reviewers should independently screen titles and abstracts, obtain and screen full texts, and record a specific reason for each full-text exclusion. Bibliographic records must be deduplicated by normalized DOI first and then by normalized title and year, retaining all source-database labels. A pilot extraction form will collect study design, site, population, educational task, application architecture, human supervision, comparison, duration, outcome instrument, effect data, and analysis limitations. Design-sensitive appraisal will distinguish randomized or quasi-experimental intervention studies, single-group pre/post studies, cross-sectional survey studies, qualitative observations, software validation, and algorithm benchmarks. Technical appraisal additionally examines training–test separation, representative audio, data leakage, class imbalance, and expert tajwid validation.

2.5. Synthesis and certainty

Because outcome measures are diverse, quantitative meta-analysis is not planned as the default. Educational assessments will be grouped only by compatible outcomes and comparison designs; machine-learning accuracy and word-error-rate estimates will be summarized separately from educational outcomes. Qualitative themes will be integrated after a design-sensitive appraisal rather than numerically pooled. Subgroup analyses are prespecified for primary versus secondary versus higher education, school versus home use, software with versus without teacher feedback, and comparison versus uncontrolled designs. The final synthesis should assess missing outcome reporting and certainty of evidence proportionately; application satisfaction is not evidence of improved retention. Publication and author-metadata discrepancies will be flagged for correction before submission.

2.6. Evidence status and Figure 1

The candidate literature map currently records 23 unique articles identified by targeted public searches of publisher or journal pages. Their DOI or article-page links, titles, and abstract-level claims were inspected for plausibility. However, there is no master export, no independent inclusion decision log, no systematic retrieval of all full texts and no completed risk-of-bias appraisal. Consequently, the records identified, screened, excluded, and included fields of a PRISMA 2020 diagram cannot responsibly be populated. Figure 1 therefore displays a stage-status diagram rather than a fictitious final selection flow. The 23 papers are candidate examples, not the final number of included studies.

Figure 1 records pending stages only; it cannot report identification, screening or inclusion counts without timestamped exports and a reconciled record ledger. Figure 1 therefore remains a protocol-stage diagram.

The appraisal plan also separates publication-record verification from claim-level validation. A correct DOI and an accessible abstract can establish the existence of a source, but they cannot confirm its randomization procedure, outcome reliability, attrition, statistical assumptions, or correspondence between a figure and a manuscript claim.

3. Theoretical Background

3.1. The educational functions of Qur'anic engagement

Digital learning studies must identify the learning construct before selecting a metric. Decoding Arabic script and producing phonemes, following tajwid, reproducing previously memorized text, developing comprehension, and sustaining motivated practice are different tasks. A single mobile application may support several of these processes, but a positive user experience does not establish equivalence among their outcomes. For example, improving a learner's ability to follow an audio model in a short exercise is not the same outcome as independently producing an accurate full surah from memory one month later. The review will therefore record both the activity supported and the instrument by which investigators claim to detect its improvement.

3.2. Human guidance and digitally mediated feedback

Pedagogically, teacher feedback provides more than a binary error notification. The teacher may explain a rule, model articulation, monitor fatigue, adapt the pace of learning, and interpret the learner's specific difficulties. Digital systems can offer additional repetitions, convenience and time-independent feedback, yet their correctness must be checked against trained expertise. The studies of speech recognition developed by (Nigar et al., 2023), (Alqadasi et al., 2023), and (Alshboul et al., 2025) illustrate three levels of technical evaluation: recognition of recited units, modeling of phoneme-duration rules, and speech-to-text word error rate. These metrics cannot substitute for blinded educator ratings of a novice's mistakes or longitudinal learning tests.

3.3. Motivation, usability and adoption

Digital memorization tools commonly combine audio playback, repetition, quizzes and gamified progression. Motivation can make more practice possible, and access can reduce practical constraints, but the mechanism is not guaranteed. When enthusiasm, perceived ease, or teacher readiness is measured, the resulting estimate must be described as such. The qualitative analysis of application reviews in (Busro & Hakim, 2025), the survey of foundation students in (Ab Alim et al., 2025), and the teacher readiness study in (Nugraha et al., 2025) illustrate complementary viewpoints. Each identifies conditions necessary for use without establishing independent, long-term learning effectiveness.

3.4. Evidence hierarchy and digital fairness

Figure 2 proposes a construct map linking a particular tool and feedback architecture to educational versus implementation outcomes. Figure 3 distinguishes technical artifact validation, usability, short-term learner improvement and generalizable educational impact. These are an analytic scaffold, not a verified causal model. The framework gives special attention to accessibility across connectivity conditions and abilities, diversity of ages and voices in training data,

and access to qualified human reciters. Overclaiming success can be as damaging as ignoring promising tools: a speech recognizer may perform well on highly controlled audio but struggle with inexperienced speakers, while a small pre/post score gain may reflect additional supervised practice rather than the mobile interface alone.

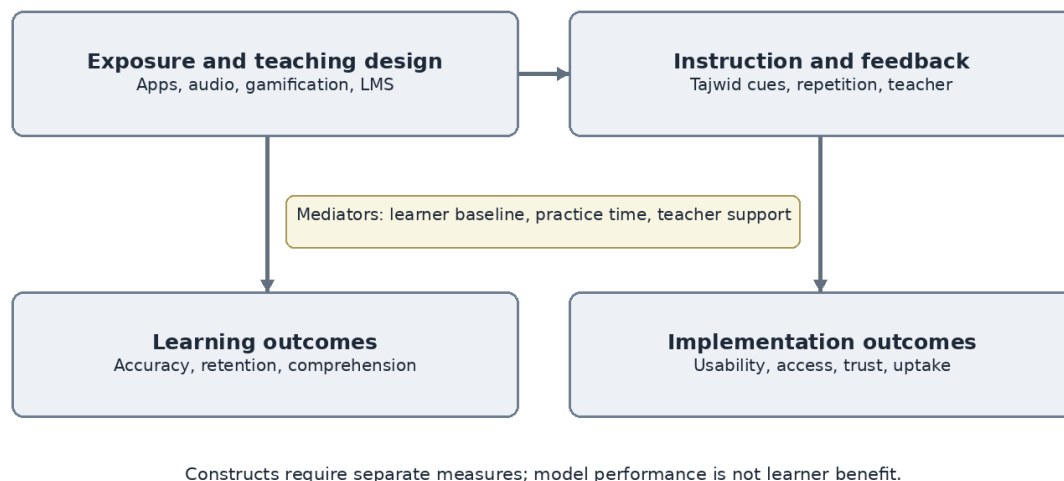


Figure 2. Construct map separating instructional processes, educational outcomes and implementation measures. Conceptual, not causal.

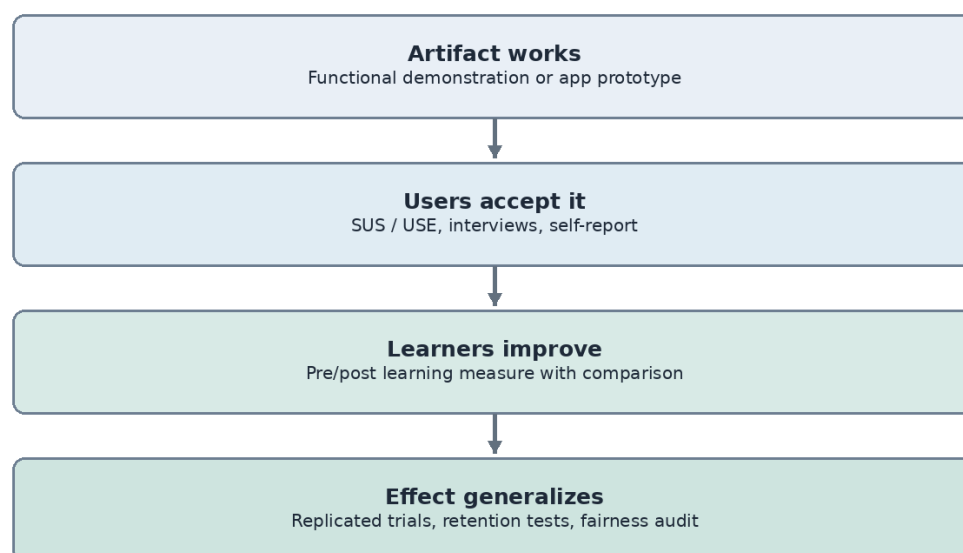


Figure 3. Evidence hierarchy for educational technology claims. Assessment levels are illustrative, not formal risk-of-bias scores.

4. Provisional Evidence Mapping

The following four strands summarize 23 candidate primary articles located from individually identified, publicly accessible journal records. They are not the results

of a completed systematic database search. Table entries are based on the source's public abstract or publisher page, with full-text methodological judgments outstanding. Frequencies, denominators, and quality rankings should not be inferred from this deliberately non-exhaustive group.

Outcome-specific appraisal matrix for digital Qur'anic learning

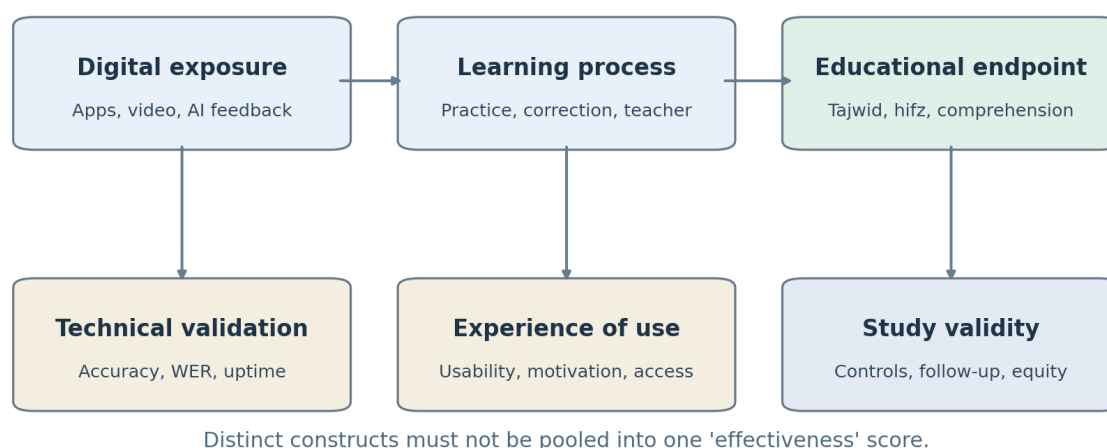


Figure 4. Outcome-specific appraisal matrix. Schematic prepared for this review; the arrows do not represent tested causal effects.

4.1. Reading, recitation and comprehension outcomes

Table 1 records five studies focusing on reading or recitation learning through digital delivery. The Bestin application used development and expert validation alongside a small pupil trial (Aeni et al., 2026), whereas (Hasjanah et al., 2022) examined online memorization outcomes in the disrupted context of remote learning. A quantitative study of video-based materials reported a positive association with reading understanding but did not establish randomized treatment effects (Salsabila & Marlina, 2024). Descriptive classroom work with projectors and YouTube emphasizes how instructional materials are delivered and supported by teachers (Sutarmi et al., 2025). A school-level pre/post examination of the NU Online digital Qur'an reported improved recitation scores among 50 learners, from 74.88 to 79.10, but lacked a parallel comparison group (Afiatna et al., 2025). In consequence, improvements reported by the studies differ in instrument, learner, exposure and ability to rule out alternative explanations. Table 1 should be read as a design-oriented preliminary inventory, not an estimate of pooled educational effects.

Table 1 should also be read alongside Figure 4: reading outcomes such as articulation, rule application, fluency and comprehension require explicitly distinct tests. In particular, the reported gain in Afiatna et al. (2025) represents uncontrolled change on the study's oral-recitation assessment, not a controlled treatment effect.

Table 1. Digital Qur'anic reading and recitation learning: candidate study extraction.

Study / source	Source-supported contribution	Design or inference limitation
(Aeni et al., 2026) Al Ibtida: Jurnal Pendidikan Guru MI	ADDIE app-development and expert-validation study with a trial involving 24 Grade 4 pupils; positive usability and reading-related evaluation reported.	Development/ evaluation
(Hasjanah et al., 2022) SOSIOEDUKASI	Investigated online learning and Qur'anic memorization outcomes during pandemic disruption in one school context.	Observational educational study
(Salsabila & Marlina, 2024) Intiqad: Jurnal Agama dan Pendidikan Islam	Quantitative study of YouTube-based instruction and pupils' reported comprehension/reading understanding.	Quantitative association
(Sutarmi et al., 2025) Al Qodiri: Jurnal Pendidikan, Sosial dan Keagamaan	Described projector and YouTube-supported digital tilawah instruction and classroom implementation.	Qualitative implementation study
(Afiatna et al., 2025) PeTeKa	A one-group pre/post study of 50 Grade 7 pupils reported mean recitation scores increasing from 74.88 to 79.10.	Pre-post study

4.2. Mobile memorization systems and online tahfiz

Six examples concern memorization tools and online practice (Table 2). An Android game based on the takrir method reported a mean pre/post score rise in a development evaluation (Djainudin et al., 2025); by contrast, application-store reviews illuminated features and reliability problems, rather than measured retention (Busro & Hakim, 2025). Field interviews revealed concerns about staff competence, infrastructure and curriculum alignment (Asyari et al., 2025), while online tahfiz interviews investigated cognitive, emotional and behavioral dimensions of engagement (Sufihara & A'isah, 2025). A questionnaire-based evaluation addressed the motivational correlates of an iOS memorization application (Hamidah & Yulius, 2024), and an Android-development report described application construction without establishing sustained learning effectiveness (Herfandi & Fitria, 2023). Table 2 shows why technical availability, user motivation and retention outcomes must be coded separately. The included examples do not establish whether a learner retains a passage after weeks or months, or whether gains depend on additional teacher contact.

A second reading of Table 2 shows a crucial temporal evidence gap. Immediate practice scores, reviewer comments and motivational questionnaires cannot substitute for delayed, independently scored recall. The distinction is central to interpretation of Djainudin et al. (2025), Hamidah and Yulius (2024), and Busro and Hakim (2025).

Table 2. Memorization applications and online tahfiz: candidate study extraction.

Study / source	Source-supported contribution	Design or inference limitation
(Djainudin et al., 2025) Jurnal Prima Edukasia	ADDIE educational game with takrir repetition; the public abstract reports mean pretest 65 and posttest 85.	Design and development with pre/post assessment
(Busro & Hakim, 2025) Ulul Albab: Jurnal Studi Islam	Feature observation and Play Store review analysis suggested benefits for engagement while revealing crashes and audio problems.	Qualitative app/review analysis
(Asyari et al., 2025) Indonesian Journal of Progressive Pedagogy	Interviews, observations and document review identified infrastructure, teacher skills and curricular barriers.	Qualitative field study
(Sufihara & A'isah, 2025) Tadrib: Jurnal Pendidikan Agama Islam	Interviews, observation and documentation explored cognitive, affective and behavioral engagement in online tahfidz.	Qualitative field study
(Hamidah & Yulius, 2024) MULTINETICS	MSLQ-based analysis reported that app-related factors account for a portion of variation in reported motivation (26.2% in abstract).	App development / questionnaire
(Herfandi & Fitria, 2023) POSITIF: Jurnal Sistem dan Teknologi Informasi	Described Android application development using observations, interviews and literature study.	Software development study

4.3. AI-assisted recitation, feedback and acoustic validation

Six candidates address automated audio recognition or AI-supported feedback (Table 3). An MFCC–LSTM speech system reported a high test accuracy for recognizing Qur'anic recitation in its chosen experimental setup (Nigar et al., 2023). A duration-sensitive model instead evaluated phoneme categories and rules relevant to tajwid (Alqadasi et al., 2023). A later study benchmarked Whisper and DeepSpeech using the QRFAM recitation dataset and evaluated word-error rate; it explicitly treated full tajwid assessment as outside its measured task (Alshboul et al., 2025). App-development research extends automated feedback into a mobile memorization interface (Khoiriyah et al., 2025), while a Malaysian survey documents the discrepancy between tool availability and real-world use among non-Islamic-studies students (Ab Alim et al., 2025). An additional quasi-experimental school article examines adaptive tools for tahfiz but requires full-text appraisal of randomization, comparability and analysis (Fauzan, 2025). Collectively these studies illustrate feasibility, design, acceptance and trial evidence—not a single standardized “AI effectiveness” outcome.

Table 3 distinguishes phoneme classification, word recognition and pedagogical feedback. A system may report 97.7% classification accuracy (Nigar et al., 2023) without validating error correction against blinded tajwid instructors or demonstrating learning transfer among novices. Figure 4 preserves the difference between technical validation and a learner endpoint.

Table 3. AI audio processing and pronunciation feedback: candidate study extraction.

Study / source	Source-supported contribution	Design or inference limitation
(Nigar et al., 2023) Applied Computational Intelligence and Soft Computing	MFCC and LSTM-RNN speech system distinguishes correct recitation at alphabet, word and verse levels; author-reported accuracy 97.7%.	Algorithm development and testing
(Alqadasi et al., 2023) Electronics	Rule-based phoneme duration modelling evaluated different madd categories on recitations from 30 participants.	Signal-processing classification
(Alshboul et al., 2025) Applied Sciences	Compared Whisper with DeepSpeech on QRFAM audio recordings; Word Error Rate was used for recognition, not full tajwid grading.	Computational benchmark
(Khoiriyah et al., 2025) Journal International Inspire Education Technology	Described a mobile AI-assisted feedback application and reported learner evaluation in a quasi-experimental component.	Design and educational assessment
(Ab Alim et al., 2025) QURANICA: International Journal of Quranic Research	Survey of 205 students found 50.7% never used the apps and 9.8% used them daily; concerns concerned teacher oversight and pronunciation feedback.	Cross-sectional survey
(Fauzan, 2025) Aslim: Journal of Education and Islamic Studies	Quasi-experimental comparison across 120 school pupils described speech-recognition and adaptive-feedback applications.	Quasi-experimental school study

4.4. Adoption, motivation, distraction and institutional conditions

Table 4 brings together six implementation-focused studies. In a mixed-method analysis, social-media exposure showed no statistically significant association with memorization levels, although qualitative respondents described focus problems (Abdelazeez et al., 2024). Teacher readiness research identified a gap between technical capability and pedagogical or institutional support (Nugraha et al., 2025). Survey data on mobile application performance and satisfaction reflect learner perceptions and cannot by themselves validate objective attainment (Hakimi et al.,

2024). Gamification analysis reported divergent directions for different mechanics and serious measurement-quality cautions (Amin et al., 2023). Two system development projects focused on prototyping and on screen-time monitoring with usability measures, respectively (Abigail & Haryono, 2025) (Abdulhafidz et al., 2025). The findings justify studying implementation barriers and possible adverse effects, but they do not determine which platform produces the strongest Qur'anic learning gains. Table 4 maps the diversity of constructs that a completed review must handle without unjustified aggregation.

The interpretation of Table 4 is necessarily conditional. The null quantitative association reported by Abdelazeez et al. (2024) should be retained alongside the interview themes, and the readiness percentages reported by Nugraha et al. (2025) remain institutional indicators rather than estimates of achievement gain.

Table 4. Adoption, gamification, usability and distraction: candidate study extraction.

Study / source	Source-supported contribution	Design or inference limitation
(Abdelazeez et al., 2024) Khalifa: Journal of Islamic Education	Survey analysis found no statistically significant relationship between social-media time and memorization level while interviews suggested focus-related challenges.	Mixed methods observational
(Nugraha et al., 2025) TARBIYA: Journal of Education in Muslim Society	Survey of 30 teachers and interviews with five found weaker pedagogical than technical readiness.	Convergent mixed methods
(Hakimi et al., 2024) Journal of Digital Learning and Distance Education	Survey elicited students' opinions on mobile app use, satisfaction and perceived gains.	Cross-sectional survey
(Amin et al., 2023) Cognoscere: Jurnal Komunikasi dan Media Pendidikan	Analysis of 250 students found mixed relationships between game mechanics, engagement and reported learning; measurement shortcomings disclosed.	Quantitative modeling
(Abigail & Haryono, 2025) International Journal of Informatics and Computation	Prototyped Android-based Quran instruction and conducted practitioner user-acceptance evaluation.	Design and usability
(Abdulhafidz et al., 2025) Jurnal Teknik Informatika (Jutif)	Developed and evaluated an app intended to redirect screen time toward Quran reading, using usability measures.	Software development/usability

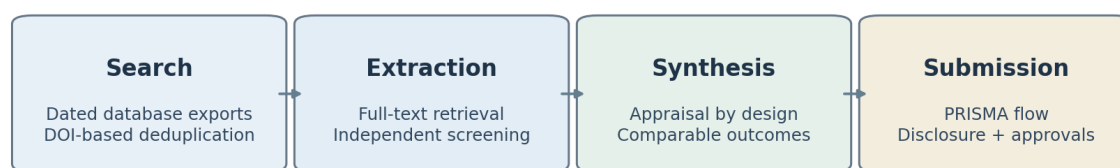
5. Discussion

The cross-strand evidence map suggests that the principal obstacle in evaluating digital Qur'anic learning is conceptual rather than technological. Researchers use effectiveness to describe different targets: functioning applications, favorable users, short-term memory scores, speech-to-text accuracy, and authentic religious-pedagogical development. Such findings can be mutually informative, but they do not form a single dependent variable. Figure 2 separates tool exposure and the instructional feedback process from learning and implementation measures; Figure 3 further distinguishes functional readiness from justified educational claims. The practical value of these distinctions is substantial. A platform can be reliable, popular and engaging yet have little demonstrated impact on retention, while a teacher-guided program can be effective even if its interface is unremarkable. A completed review must resist treating technical sophistication as a surrogate outcome.

Comparative claims need particularly careful attention to research design. Most of the mapped applications have been examined in single schools or institutions, with samples assembled for convenience, development testing or local program evaluation. Where pre/post measures are available, the improvement may reflect repeated testing, extra instructional attention, maturation, parent involvement or differences in assessment content. The rise in pre/post mean memorization scores reported for Qur'an Whiz, for example, is relevant to the local application evaluation but does not establish its advantage over equivalent teacher-led repetition (Djainudin et al., 2025). By contrast, observational work can illuminate real constraints even though it cannot identify intervention effects. Excluding all qualitative studies would lose the teacher, institutional and accessibility conditions on which effectiveness plausibly depends.

Speech technologies create a second comparison challenge: prediction measures and educational outcomes have different units of analysis. A model can classify a sound correctly on a pre-labeled dataset without determining whether a novice has internalized a tajwid rule. Training data may overrepresent fluent adult voices and omit dialectal, age-related or novice articulatory diversity. Recognition and correction should also be assessed separately: the model may transcribe correctly but provide feedback that is unclear, overconfident or insensitive to the teaching context. Future trials should retain model version, noise conditions, sampling of speakers, reference transcriptions and expert adjudication protocols. Studies based on highly controlled audio, such as the investigations of recitation recognition in (Nigar et al., 2023) and (Alshboul et al., 2025), make a valuable engineering contribution without settling these downstream educational questions.

Publication readiness requires a traceable evidence chain



Missing exports and reviewer decisions cannot be reconstructed from a selected article list.

Figure 5. Evidence-to-submission audit chain. A conceptual workflow identifying uncompleted verification requirements.

A stronger evidence chain would test both recognition validity and instructional usefulness: independent expert raters should assess the presence of genuine tajwid errors, model outputs should be compared with those ratings, and an educational evaluation should determine whether learners improve after receiving the advice. Figure 5 sets out the corresponding research and reporting gates.

The learner's situation is equally important. Qur'anic education is often relational, involving teacher modeling, correction and continuity of practice. Mobile tools can extend practice, but their convenience is shaped by access to stable connectivity, suitable devices, Arabic literacy, quiet learning spaces and technical support. Institutional competence can determine whether an app becomes part of a learning sequence or an isolated novelty. Studies of digital memorization in field settings and teacher readiness suggest why training and organizational support belong inside the causal narrative of an intervention rather than in a peripheral paragraph (Asyari et al., 2025) (Nugraha et al., 2025). At the same time, the relatively low everyday application use reported in a student survey suggests that supply-side development alone cannot be equated with uptake (Ab Alim et al., 2025).

Gamification offers a concrete example of uncertain mechanisms. Repetition, progress indicators, badges and leaderboards may influence practice, social comparison and persistence in different ways. In the preliminary set, a gamification study reported mixed associations between game mechanics and learning, together with material concerns about measurement reliability (Amin et al., 2023). No broad claim that gamification improves Qur'anic learning should be drawn from such an analysis. A useful future trial would preregister a theory of change, randomize comparable learning groups when feasible, report engagement and performance separately, and measure retained recitation after an appropriate follow-up interval. Such a design would also examine whether competitive mechanics are acceptable to educators and learners in different settings.

A balanced review cannot ignore contradictory or null findings. Technology-related distractions are plausible, but evidence of concern is not evidence of a statistically reliable adverse effect. The mixed-method distraction study reported no significant association in its quantitative analysis, even while interview accounts described concentration difficulties (Abdelazeez et al., 2024). Differences may reflect

measurement error, sample size, device practices or contextual confounding rather than an intrinsic conflict between statistical and experiential evidence. A comprehensive synthesis should extract the exact exposure definition and statistical estimate, and present narratives as explanations to investigate, not substitutes for the reported analysis. Likewise, satisfaction and acceptance findings are informative for design but cannot resolve whether memorization became more accurate.

A future systematic review should also address inequalities within the evidence base. Searches restricted to English-language education journals could miss locally published primary research in Bahasa Indonesia, Arabic or Malay; the availability of English abstracts should not be confused with availability of eligible full texts. Computational datasets may favor adult expert reciters, while educational trials may center on students in digitally equipped institutions. Reporting of disability access, device ownership, rural connectivity, teacher qualifications, and oral-recitation reference standards should become standard extraction fields. Interventions should be evaluated for whether their automated advice complements or displaces trusted human guidance, and for whether they provide clear pathways to human correction when uncertain.

The present manuscript has significant limitations. Public publisher searches were used to locate a purposive set of 23 candidate primary articles; no auditable Scopus, Web of Science or ERIC exports were available. There has been no double screening, full-text duplicate appraisal or reconciled exclusion log. Some bibliographic particulars, publication-language status and study-specific measures remain to be checked against complete papers. The preliminary set may overrepresent studies with visible journal pages and English abstracts and underrepresent work archived elsewhere. For this reason, its four tables are an evidence map intended to support question refinement, not final systematic review results. Claims about prevalence, global effect size, study-quality distribution or publication bias would be premature.

The incomplete status of evidence screening is the principal limitation of this submission-stage revision. The 23 candidate records are not denominators for a systematic review and should not be represented as PRISMA-included studies. As Figure 5 indicates, completing the bibliographic export, reconciliation, and full-text appraisal is a prerequisite—not an optional cosmetic step—before submission as an evidence synthesis.

A rigorous next phase would create a dated corpus, implement the full searches and deduplicate records, then use design-appropriate appraisal to decide which comparisons merit quantitative summaries. Publication-ready reporting should make clear whether the objective is to assess intervention effects, explain implementation, validate computational feedback or integrate these streams through a mixed-method synthesis. It should not obscure weak evidence behind inflated reference counts. Practical recommendations should remain proportional: developers should document speech-training corpora and feedback limitations; teachers should assess tools against explicit curricular objectives; researchers should conduct controlled outcome and retention tests; and reviewers should report implementation constraints alongside benefits. These actions can support innovation while maintaining the epistemic standards expected of research on a sacred educational text.

6. Conclusion

Digital tools can expand access to Qur'anic instruction, offer repeated listening and memorization practice, support formative feedback, and generate new forms of classroom and home-based interaction. The preliminary articles mapped here also document risks of conceptual overreach: tests of speech-recognition systems, software usability assessments, learner satisfaction surveys, and education outcomes are distinct evidential claims. The reviewed examples suggest that short-term benefits are possible in specific contexts, but their heterogeneous designs do not justify a universal statement about the superiority of digital over teacher-led Qur'anic learning.

The central contribution of this systematic-review protocol is a framework for evaluating technology as part of a pedagogical process, not as an inherently effective intervention. A completed review must distinguish decoding accuracy, tajwid proficiency, memorization retention, comprehension, motivation and accessibility, and must account for teacher support, educational level and the nature of comparator conditions. Computational measures should be checked against expert judgments and downstream learner outcomes. Because an exhaustive corpus, independent screening and full-text appraisal remain outstanding, the 23 mapped papers constitute preliminary candidates rather than a finalized included set. Future studies should favor transparent comparisons, longer follow-up, reproducible evaluation datasets, and equitable access. These steps are essential if decisions about digital Qur'anic learning are to reflect the quality of evidence rather than technological enthusiasm.

In practical terms, the next research cycle should preregister the protocol, archive deduplicated records and adjudicated eligibility decisions, repeat the analysis after excluding high-risk-of-bias studies, and explicitly compare findings with and without records with uncertain full-text eligibility.

References

- Ab Alim, N. N. A., Wahid, K., Abd Zamani, S. N. H., & Adenan, F. (2025). Artificial intelligence as a support tool for Quran memorization: An exploratory study on the potential use of AI applications among non-Islamic studies students. *QURANICA: International Journal of Quranic Research*, 17(2). <https://doi.org/10.22452/quranica.vol17no2.7>
- Abdelazeez, S., Obeid, A. R., Lukman, M., Alhussian, H., Aziz, N., & Patel, A. (2024). Digital distraction in Quranic education: A mixed methods approach. *Khalifa: Journal of Islamic Education*, 8(1), 11–26. <https://doi.org/10.24036/kjie.v8i1.405>
- Abdulhafidz, Y., Zaky, U., & Admojo, F. T. (2025). Development of mobile Quran app with screen time monitoring using DRM, Agile, and SUS-USE testing. *Jurnal Teknik Informatika (Jutif)*, 6(6). <https://doi.org/10.52436/1.jutif.2025.6.6.5398>
- Abigail, M. I., & Haryono, K. (2025). Building an effective Quran learning system with prototyping model. *International Journal of Informatics and Computation*, 7(2), 355–365. <https://doi.org/10.35842/ijicom.v7i2.138>
- Aeni, A. N., Salmah, H. S., Agistie, D. S., Nurainun, S. R., & Muhammad, N. (2026). Development of the Bestin Android App to Improve the Reading Qur'an of

- Students Elementary Schools. *Al Ibtida: Jurnal Pendidikan Guru MI*, 13(1). <https://doi.org/10.24235/al.ibtida.snj.v13i1.24592>
- Afiatna, F., Fauziyah, N. L., & Hidayatulloh, M. A. (2025). Efektivitas penggunaan aplikasi digital Al-Qur'an (NU Online) terhadap peningkatan kemampuan tilawah siswa SMP Islam An-Nadwah. *PeTeKa*, 8(4). <https://jurnal.um-tapsel.ac.id/index.php/ptk/article/view/23346/0>
- Alqadasi, A. M. A., Sunar, M. S., Turaev, S., Abdulghafor, R., Hj Salam, M. S., Alashbi, A. A. S., Salem, A. A., & Ali, M. A. H. (2023). Rule-based embedded HMMs phoneme classification to improve Qur'anic recitation recognition. *Electronics*, 12(1), Article 176. <https://doi.org/10.3390/electronics12010176>
- Alshboul, M., Al Muaitah, A. R., Al-Issa, S., & Al-Ayyoub, M. (2025). Enhanced neural speech recognition of Quranic recitations via a large audio model. *Applied Sciences*, 15(17), Article 9521. <https://doi.org/10.3390/app15179521>
- Amin, S., Nurfadilla, A., Rasman, Puspitasari, A., & Baskautshar, N. (2023). Gamification in Qur'anic learning: Evidence from Muslim students in hybrid classrooms. *Cognoscere: Jurnal Komunikasi dan Media Pendidikan*, 1(2), 60–67. <https://doi.org/10.61292/cognoscere.285>
- Asyari, A., Hamami, T., & Sukiman, S. (2025). The role of digital applications in enhancing Qur'an memorization learning. *Indonesian Journal of Progressive Pedagogy*, 1(2), 116–127. <https://doi.org/10.61987/ijpp.v1i2.1265>
- Busro, B., & Hakim, R. (2025). Click, learn, remember: Digital tools for memorizing the Quran. *Ulul Albab: Jurnal Studi Islam*, 26(2), 282–304. <https://doi.org/10.18860/ua.v26i2.32440>
- Djainudin, H., Saputra, A. A., Rahmat, N. N., & Aprilio, T. (2025). Qur'an Whiz: Developing an Android-based application to enhance Qur'an memorization skills for elementary school students. *Jurnal Prima Edukasia*, 13(1), 85–97. <https://doi.org/10.21831/jpe.v13i1.80349>
- Fauzan, T. S. (2025). Implementation of artificial intelligence in tahfidz Al-Qur'an learning at Madrasah Tsanawiyah: A study on the effectiveness of machine learning-based applications. *Aslim: Journal of Education and Islamic Studies*, 2(3), 209–215. <https://doi.org/10.63738/aslim.v2i3.42>
- Hakimi, M., Akrami, M., Ahrari, M., Akrami, K., & Akrami, F. (2024). The impact of mobile applications on Quran education: A survey of student performance and satisfaction. *Journal of Digital Learning and Distance Education*, 2(9), 722–736. <https://doi.org/10.56778/jdlde.v2i8.220>
- Hamidah, D. N. H., & Yulius, R. (2024). Analysis of the influence of iOS based Al-Quran Juz 30 memorizing applications on user motivation using the MSLQ model. *MULTINETICS*, 10(1), 60–72. <https://doi.org/10.32722/multinetics.v10i1.6672>
- Hasjanah, K., Kurniawan, N. B., & Siagian, T. H. (2022). The impact of online learning implementation on student's learning outcomes in Al Quran memorization. *SOSIOEDUKASI*, 11(1), 89–99. <https://doi.org/10.36526/sosioedukasi.v11i1.1927>

- Herfandi, & Fitria, A. (2023). Development of an Android-based Quran memorization application. *POSITIF: Jurnal Sistem dan Teknologi Informasi*, 8(2), 107–113. <https://doi.org/10.31961/positif.v8i2.1486>
- Khoiriyah, R., Hasanah, E. S., Al Hidayah, A., & Susrianingsih, S. (2025). Innovation in Islamic educational technology: Development of a mobile application for Qur'anic memorization (Tahfidz) with AI-based feedback. *Journal International Inspire Education Technology*, 4(2), 180–197. <https://doi.org/10.55849/jiiet.v4i2.1021>
- Nigar, N., Wajid, A., Ajagbe, S. A., & Adigun, M. O. (2023). An intelligent framework based on deep learning for online Quran learning during pandemic. *Applied Computational Intelligence and Soft Computing*, 2023, Article 5541699. <https://doi.org/10.1155/2023/5541699>
- Nugraha, N. A. A., Firmansyah, M. I., & Nugraha, R. H. (2025). AI-gamification in Qur'anic learning: Assessing Indonesian Islamic religious education teachers' readiness for Generation Alpha. *TARBIYA: Journal of Education in Muslim Society*, 12(1), 141–162. <https://doi.org/10.15408/tjems.v12i1.50479>
- Salsabila, A., & Marlina, Y. (2024). The impact of IT-based learning media use on improving understanding of reading the Qur'an. *Intiqad: Jurnal Agama dan Pendidikan Islam*, 16(1), 48–56. <https://doi.org/10.30596/19848>
- Sufihara, I., & A'isah, A. (2025). Memorizing the Qur'an in the digital era: Santri learning engagement through online tahfidz programs. *Tadrib: Jurnal Pendidikan Agama Islam*, 11(1), 283–292. <https://doi.org/10.19109/tadrib.v1i1.30680>
- Sutarmi, Jannah, S. R., & Setiawan, A. (2025). Implementasi pembelajaran Al-Qur'an berbasis digital dalam meningkatkan literasi dan motivasi belajar siswa MTs Darussalam Banjar Negeri Kecamatan Natar Lampung Selatan. *Al Qodiri: Jurnal Pendidikan, Sosial dan Keagamaan*, 23(1), 138–155. <https://doi.org/10.53515/qodiri.2025.23.1.138-155>
- Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*, 368, l6890. <https://doi.org/10.1136/bmj.l6890>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: An extension to the PRISMA Statement for reporting literature searches in systematic reviews. *Systematic Reviews*, 10, Article 39. <https://doi.org/10.1186/s13643-020-01542-z>