

MNEMONIC STRATEGIES IN ARABIC VOCABULARY AND GRAMMAR LEARNING: A SYSTEMATIC LITERATURE REVIEW

Hikmatul Lailaa¹ , Rasha Kadry Soliman² , Ammar Zainuddin³ 

¹Pendidikan Bahasa Arab, Universitas Negeri Makassar, Indonesia

²University of Leeds, United Kingdom

³Universitas KH. Abdul Chalim, Indonesia

E-mail of Authors: ¹hikmatul.lailaa@unm.ac.id, ²R.K.Soliman@leeds.ac.uk ³ammar@uac.ac.id

Received: 29-09-2026

Revised: 5-10-2026

Accepted: 6-10-2026

Published: 16-10-2026

ABSTRACT

Mastery of Arabic vocabulary (mufradat) and grammar rules still relies on rote memorisation, which students struggle to retain, and mnemonic strategies are increasingly being tested as an alternative. However, no systematic review has mapped the techniques, outcomes, and strength of evidence in these studies. This study synthesises empirical evidence on mnemonic strategies for learning Arabic vocabulary and grammar. It employs a Systematic Literature Review following the PRISMA protocol with narrative synthesis. The online search yielded 31 unique records, of which 11 studies published between 2020 and 2026 met the inclusion criteria of this review. Seven studies addressed vocabulary, and four addressed grammar rules, mainly morphology (sharaf). The dominant techniques were rhyme or song, keyword mnemonics, and story association with visualisation. All studies with readable results reported improvement or a positive trend, although most used non-controlled or single-group designs. A quality appraisal using the Mixed Methods Appraisal Tool (MMAT) found that only two studies met most core quality criteria. Therefore, the evidence supports a positive direction of reported outcomes but is insufficient to establish mnemonic effectiveness. This review provides a map of techniques, subject domains, and design strengths, and points to the need for controlled experiments comparing techniques, delayed retention tests, and measurement of motivation.

Keywords: *mnemonic strategies, Arabic vocabulary, Arabic grammar, Arabic language learning, systematic literature review.*

المخلص:

لا يزال إتقان المفردات وقواعد اللغة العربية يعتمد على الحفظ، وهو ما يصعب استمراره في ذاكرة الطلاب، مما دفع إلى اختبار العديد من الاستراتيجيات التذكيرية كبديل. ومع ذلك، لم يتم إجراء أي مراجعة منهجية حتى الآن تحدد التقنيات والنتائج وقوة الأدلة المستمدة من تلك الدراسات. تهدف هذه الدراسة إلى تجميع الأدلة التجريبية حول استراتيجيات التذكر في تعلم المفردات وقواعد اللغة العربية. وقد استخدمت في ذلك «المراجعة المنهجية للأدبيات» وفق بروتوكول PRISMA مع التأليف السردي. أسفر البحث عبر الإنترنت عن 31 سجلًا فريدًا، واستوفت 11 دراسة نُشرت في الفترة من 2020 إلى 2026 معايير الإدراج. أظهرت النتائج أن سبع دراسات استهدفت المفردات وأربع دراسات استهدفت القواعد، لا سيما علم الصرف،

مع استخدام تقنيات سائدة تتمثل في القافية أو الأغنية، والكلمات المفتاحية التذكيرية، بالإضافة إلى الربط القصصي والتصور الذهني. أبلغت جميع الدراسات التي تم تحليل نتائجها عن تحسن أو اتجاه إيجابي، رغم أن غالبيتها اعتمدت تصاميم غير مضبوطة أو أحادية المجموعة. وقد أظهر تقييم الجودة باستخدام أداة MMAT أن دراستين فقط استوفتا معظم معايير الجودة الأساسية. وختامًا، تدعم الأدلة وجود اتجاه إيجابي في النتائج المُبلَّغ عنها، لكنها غير كافية لإثبات فعالية استراتيجيات التذكر. تساهم هذه الدراسة في تقديم خريطة للتقنيات ومجالات المواد وقوة تصميم الدراسات، كما تبرز الحاجة إلى إجراء تجارب مضبوطة تقارن بين التقنيات، واختبارات الاحتفاظ المتأخر، وقياس الدافع. الكلمات المفتاحية: استراتيجيات التذكر؛ المفردات؛ القواعد؛ تعلم اللغة العربية؛ المراجعة المنهجية للأدبيات.

INTRODUCTION

Mastery of vocabulary (mufradat) and grammatical rules (kaidah), particularly nahwu (syntax) and sharaf (morphology), is the foundation for all Arabic language skills. However, both also represent the aspects that madrasah students find most challenging because they rely heavily on rote memorisation. The practice of repeated memorisation without an efficient memory strategy causes memorised vocabulary to be quickly forgotten, so that students' literacy in Arabic texts fails to develop optimally (Mulloh, 2025a). A similar difficulty appears in morphology, where students must remember and apply patterns of verb and noun inflection, while traditional instruction tends to rely on memorisation without a strategic approach that supports retention (Asadi et al., 2023). Consequently, the learning process becomes lengthy and tiring, and teachers frequently lack alternative memory strategies grounded in cognitive research. Thus, the gap between curricular demands and students' actual achievement persists in many classrooms.

Low student engagement is often regarded as the primary cause, and many improvement efforts are directed toward motivation. Motivation to learn Arabic is influenced by internal factors such as interest and self-efficacy, as well as external factors such as the quality of instruction (Widayanti et al., 2025). Technology-based interventions, such as learning management systems, have been reported to increase motivation, but their focus lies on the management of learning rather than on the cognitive processes that determine whether vocabulary and grammatical rules are retained (Fauzi & Anindiati, 2021). Increased motivation does not necessarily result in improved retention because obstacles at the encoding and retrieval stages remain. Therefore, a strategy that directly targets memory is required.

Mnemonic strategies can address this need. Their theoretical foundation is derived from Dual Coding Theory, which posits that information encoded through both verbal and visual channels simultaneously is more easily retained and retrieved (Alibasya et al., 2024). A derivative technique, the keyword mnemonic, links a foreign word to a similar-sounding word already familiar to the student and reinforces this link with a mental image (Chiu & Hawkins, 2023). Unlike ordinary memorisation, this technique does not rely solely on repetition but builds a network of meaningful associations in the long-term memory.

Interest in mnemonics is not confined to language learning alone. Its effectiveness has been tested, for instance, on the ability to memorize physics formulas (Jusriana & Jurdah

Budiman, 2022), indicating that the idea is regarded as relevant across fields of study. In Arabic vocabulary, the application of the keyword mnemonic to junior secondary madrasah (MTs) students was tested through a quasi-experimental design with experimental and control classes (Ferdinand & Yasry, 2023a). The characteristics of Arabic, which demand mastery of both vocabulary and grammatical rules, make it worthwhile to examine this strategy from both aspects.

Empirical evidence has also begun to emerge in the domain of grammatical rules. A quasi-experiment on sharaf instruction that combined rhyme-based mnemonics with a deductive approach reported an increase in mean scores from 61.50 to 81.67, with a paired t-test showing a significant difference (Sholiha et al., 2023a). A two-group experiment on ilm sharf at an Islamic boarding school (pesantren) reported a calculated t-value of 4.5, exceeding the table t-value at both the 1% and 5% significance levels (Kadir et al., 2024a). Both findings indicate that mnemonics have been examined for both vocabulary and morphological patterns.

To verify this gap, a targeted search for prior reviews on mnemonic strategies in Arabic language learning was conducted in October 2026 using the terms ("systematic review" OR "literature review") AND ("mnemonic" OR "mnemonics") AND ("Arabic language" OR "bahasa Arab") across a general web search engine, DOAJ, and academic search engines. This search retrieved two related works: a qualitative descriptive literature study on the use of mnemonic methods in Arabic learning, which did not follow a systematic protocol such as PRISMA (Amila Sholiha & Asa Ismia Bunga Aisyahrani, 2023), and a PRISMA-based systematic review on Arabic speaking-skill strategies, which covers diverse instructional approaches but does not isolate or specifically examine mnemonic techniques (Aini et al., 2025). No existing review combines a systematic PRISMA protocol with a specific focus on mnemonic strategies applied to both Arabic vocabulary and grammatical rules. This constitutes the precise gap addressed here: synthesising empirical evidence on mnemonic techniques across both domains through an explicit and reproducible selection protocol, a combination not previously mapped systematically.

The need for such an evidence map is increasingly apparent, given the varying designs and strengths of the available primary studies. A one-group pre-experiment with 18 students recorded a score increase from 41.67 to 77.22, with an N-gain in the moderate category (Mahendra & Nasir, 2025a). In contrast, a two-group pilot experiment on numeral rules found only a positive tendency, with the author affirming that the experiment's limitations preclude a definitive conclusion (Cojanu, 2025a). Without a synthesis, teachers and curriculum developers cannot distinguish sufficiently robust findings from those that remain so preliminary.

A Systematic Literature Review following the PRISMA protocol was chosen because it allows for a transparent, replicable, and well-documented study selection process across the identification, screening, eligibility, and inclusion stages (Page et al., 2021a). The systematic review approach has also been used to map evidence in Arabic language learning, for example, on gamification for non-Arabic speakers (Almelhes, 2024a). The theoretical contribution of this study is to map the mnemonic techniques used and the strength of the evidence in two domains:

vocabulary and grammatical rules. Its practical contribution lies in providing a reference for teachers in selecting appropriate techniques while marking research gaps for subsequent primary studies.

Based on the foregoing, this study aims to synthesise empirical evidence on mnemonic strategies for Arabic vocabulary and grammatical-rule learning through a PRISMA-based Systematic Literature Review. The PICO framework applied defines the population as Arabic language learners at various educational levels, the intervention as mnemonic strategies in their various forms, the comparison as conventional instruction or the condition prior to treatment, and the outcome as vocabulary and grammatical-rule learning outcomes, with retention and motivation as supporting findings of the study. The research questions concern which mnemonic techniques have been used in Arabic vocabulary and grammar instruction in studies from 2020-2026, the extent to which the available evidence supports their effectiveness in both domains, and the strength of study designs and the remaining research gaps.

RESEARCH METHODS

This study employs a Systematic Literature Review (SLR) approach with the PRISMA protocol to ensure that the processes of identification, screening, eligibility assessment, and study inclusion are transparent and replicable (Page et al., 2021b). This approach was chosen because the aim of the study was not to test mnemonic strategies on a particular population but to map and synthesise evidence from already published primary studies. The review is guided by three questions: which mnemonic techniques are used in teaching Arabic vocabulary and grammatical rules, the extent to which the evidence supports their effectiveness in both domains, and the strength of study designs and remaining research gaps.

To ensure consistent study selection, the key terms used in the inclusion criteria were defined as follows. Mnemonic strategies refer to instructional techniques intentionally designed to support memory encoding and/or retrieval through association, imagery, verbal elaboration, or structured patterning (e.g. the keyword method, rhyme or song, story association, acronyms, and chunking), as distinct from unaided rote repetition. Grammar (*kaidah*) refers to rule-governed aspects of Arabic morphology (*sharaf*) and syntax (*nahwu*), assessed through tasks requiring the correct application, recognition, or recall of grammatical patterns (e.g., verb conjugation, word pattern/*wazn*, sentence structure), as distinct from vocabulary meaning recall. Retention refers to performance on a vocabulary or grammar measure administered after the initial learning session, whether immediately or after a delay, as reported in the original study. Most of the included studies measured only immediate post-treatment retention, a limitation that is discussed further below. Learning outcomes refer to any quantifiable or qualitatively reported change in students' vocabulary or grammar performance, knowledge, or skills attributable to the instructional intervention, as measured by the study's own instrument (test score, percentage correct, or qualitative report).

Table 1. *Inclusion and exclusion criteria*

Criterion	Inclusion	Exclusion
Publication year	2020-2026	Before 2020
Article language	Indonesian and English	Other than these
Type of study	Empirical primary studies (experimental, quasi-/pre-experimental, classroom action research, mixed-methods, qualitative field research) reporting learning outcomes, retention, or perceptions	Literature reviews, opinion pieces, unpublished theses or dissertations not published in a journal
Study subjects	Arabic language learners at all educational levels	Learners of other languages
Intervention	Mnemonic strategies or techniques (keyword method, imagery, storytelling, acronyms, chunking, rhyme or song as memory aids)	Other methods (mind mapping, flashcards, TPR, card sort, mimicry memorization)
Outcome variable	Vocabulary and/or grammatical-rule learning outcomes (syntax, morphology, grammar); retention and motivation as supporting outcomes	Primary outcome consisting of other skills (reading, writing, speaking) without vocabulary or grammar data
Accessibility	Full text accessible	Only abstract available

The search was conducted between 15 and 30 September 2026 by a single reviewer using a general web search engine (Google-indexed results), the Directory of Open Access Journals (DOAJ), and individual open-access journal portals (OJS-based national journal sites). No subscription databases (e.g. Scopus, Web of Science) or institutional repositories requiring access credentials were searched, which is acknowledged as a limitation of coverage. The complete set of search terms used, applied iteratively and in combination, was *mnemonik*, *mnemonic*, *"jembatan keledai"*, *"keyword method"*, *akronim*, *rima*, *imagery*, *lagu/song*, combined with *bahasa Arab*, *Arabic*, and *mufradat*, *kosakata*, *vocabulary*, *nahwu*, *sharaf*, *qawaid*, *grammar*, *morphology*, *hijaiyah*, *kitabah*, *qiraah*, *maharah*. The two representative Boolean strings were: for Indonesian-language sources, (*"mnemonik" OR "mnemonic" OR "jembatan keledai" OR "keyword method" OR akronim OR rima*) AND (*"bahasa Arab"*) AND (*mufradat OR kosakata OR nahwu OR sharaf OR qawaid*); for English-language sources, (*mnemonic* OR "keyword method" OR imagery OR acronym*) AND *Arabic* AND (*vocabulary OR grammar OR morphology*). Because preliminary searches returned few mnemonic-specific Arabic studies, the keyword set was progressively broadened across successive searches rather than being applied as a single fixed query, which limits exact reproducibility and is noted as a further limitation.

Selection proceeded through four stages in line with the PRISMA flow: record identification, screening of titles and abstracts, full-text eligibility assessment based on Table 1, and determination of the included studies.

Data from each study that passed selection were extracted into a structured table with the following fields: author and year, country or context, study design, population and sample, type of mnemonic technique, subject-matter domain (vocabulary or grammatical rules), and key findings. Because the designs of the primary studies varied, ranging from experiments with a control class, one-group pre-experiments, classroom action research (PTK), to qualitative descriptive studies, the designs were categorised descriptively to mark the strength of evidence. The thematic categories subsequently used to group mnemonic techniques in the synthesis (sound-based, imagery-based, and narrative-based) were derived inductively by the authors during this extraction process, based on recurring patterns across the included studies rather than from a predefined coding framework. The synthesis was carried out narratively, grouped according to subject-matter domain and design strength, without meta-analysis owing to the differences in design and measurement instruments across studies.

Because the included studies employed heterogeneous designs (qualitative, nonrandomised experimental, and quantitative descriptive), the methodological quality was appraised using the Mixed Methods Appraisal Tool (MMAT) version 2018 (Hong et al., 2018), which provides five core criteria per design category rather than a single uniform checklist applicable to all designs. Each study was rated as meeting or not meeting its category's five criteria based on the information reported in the original publication; criteria that could not be verified from the available text were marked accordingly. This appraisal does not replace a full systematic risk-of-bias assessment of every methodological element; however, it provides a transparent, design-appropriate indicator of evidence strength that complements the narrative synthesis. The results are shown in Table 5.

RESULTS AND DISCUSSION

Study Selection

The search yielded 31 unique records [insert references]. After removing one duplicate, 30 records were screened based on their titles and abstracts, and 18 were assessed at the full-text stage. Eleven studies met all the criteria listed in Table 1 and were further analysed. The selection flow is illustrated in Figure 1. These figures are internally consistent: 31 identified records minus 1 duplicate yields 30 screened; 30 screened minus 12 excluded at screening yields 18 assessed for eligibility; and 18 minus 7 excluded at eligibility yields the 11 included studies.

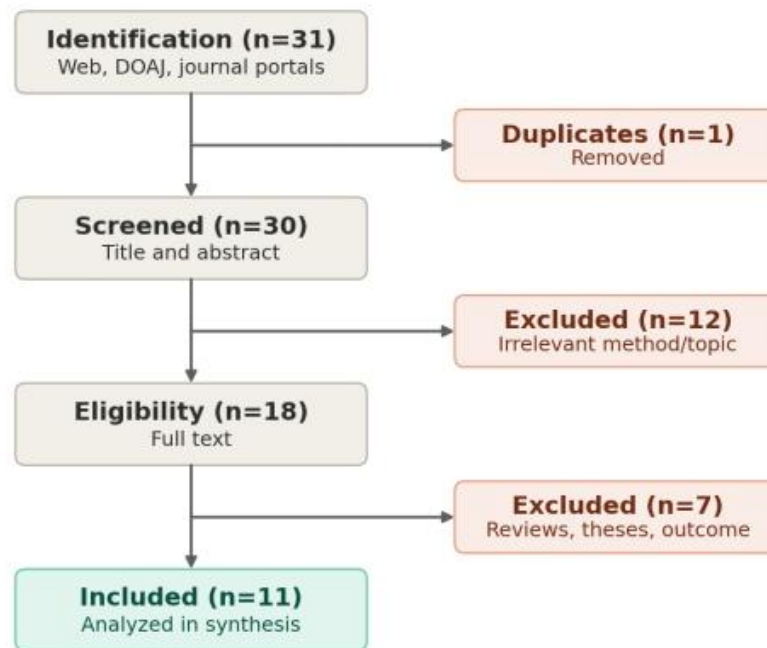


Figure 1. PRISMA study selection flow diagram

At the screening stage, 12 records were excluded: ten because their method, language, or field was not relevant (mind mapping, flashcards, TPR, card sort, word cards, mimicry memorisation, two articles, a physics study, an English-language study, and an EFL study) and two because they were systematic reviews that did not focus on mnemonics. At the eligibility stage, seven records were excluded: four literature or narrative reviews, two theses or dissertations, and one study whose primary outcome was reading skill. One duplicate record was removed prior to the screening.

Study Characteristics

The 11 studies were published between 2020 and 2026, distributed as follows: 1 (2020), 2 (2021), 2 (2023), 1 (2024), 4 (2025), and 1 (2026). Almost all were set in madrasahs or Islamic boarding schools (pesantren) in Indonesia, with one study spanning both Indonesia and Malaysia. The identified sample sizes were small, ranging from 17 to 90. The characteristics of each study are presented in Table 2, and the distribution of designs by subject matter domain is presented in Table 3.

Table 2. Characteristics and key findings of the included studies

No	Author (year)	Context/ sample	Design	Mnemonic technique	Domain	Key finding
1	Firdaus & Hafidah (2020)	MA Nurul Jadid, female students and teacher	Qualitative descriptive (observation, interviews)	Story association, rhyme	Vocabulary	Learning proceeded smoothly and female students were more enthusiastic; concluded effective without score data
2	Nasir & Basit (2021)	Experimental and control classes	Quasi-experiment with control class	Keyword mnemonic	Vocabulary	Pretest 66.07 (experimental) and 62.18 (control); experimental posttest mean 85.59
3	Imron & Fajriyah (2021)	MI Baitul Huda, 22 respondents	Experiment with control class	Singing method (song)	Vocabulary	Experimental class mean 90.90; calculated r of 0.582 exceeding the table r at the 1% level
4	Ferdinand & Yasry (2023)	MTs Plus Darul Hufaz; class VII B (experimental), VII A (control)	Quasi-experiment	Keyword mnemonic	Vocabulary	Score increased from 49.00 to 86.00, a rise of 37 points following implementation
5	Sholiha et al. (2023)	Interdisciplinary Islamic Studies students, UNU Yogyakarta	One-group quasi-experiment	Rhyme (sholawat badar) + deductive approach	Grammar (sharaf)	61.50 -> 81.67; paired t-test significant (0.000)
6	Kadir et al. (2024)	Al-Falah Islamic Boarding School, West Limboto	Two-group experiment	Mnemonic method	Grammar (sharaf)	Calculated t of 4.5, exceeding the table t of 2.44 (1%) and 1.69 (5%)
7	Mahendra & Nasir (2025)	MTs Manba'ul Ulum, 18 students	One-group pre-experiment	Pictures, storytelling, song, word grouping	Vocabulary	41.67 -> 77.22; N-Gain 0.6288 (moderate)
8	Baharun et al. (2025)	90 students (30 Indonesia, 60 Malaysia), 16-18 years	Quasi-experiment pre-post + interviews, no control group	Mnemonics, visualization, and other strategies	Vocabulary (50 words)	Mean 58.3 -> 81.4
9	Hanifansyah (2025)	30 students, 16-18 years	Qualitative descriptive	Acronyms, visualization, storytelling + muscle memory	Grammar (sharaf)	Pattern retention 45% -> 82%; comprehension 40% -> 78% (descriptive data)
10	Cojanu (2025)	Not identified	Two-group pilot	Mnemonic phrases	Grammar (numerals)	Positive trend; author states limitations

			experiment			preclude a definitive conclusion
11	Awaliyah et al. (2026)	SMPIT Makassar Islamic School, 17 students	Two-cycle classroom action research	Story association (Cycle I), rhyme (Cycle II)	Vocabulary	84.94 -> 88.41; students below the minimum mastery criterion (KKM) 8 -> 0

Table 3. *Distribution of studies by design and domain*

Design	Vocabulary	Grammar	Total
Experiment with control class	3	2 (one pilot)	5
One group without control	2	1	3
Classroom action research	1	0	1
Qualitative descriptive	1	1	2
Total	7	4	11

Quality Appraisal of Included Studies

As described in the Research Methods section, the methodological quality was appraised using the MMAT (Hong et al., 2018). Table 5 shows that only Kadir et al. (2024) and Baharun et al. (2025) met most of their category's core criteria, while the remaining studies, particularly Cojanu (2025) and the two qualitative studies, left multiple criteria unmet or unverified. This appraisal corroborates the design-based synthesis presented below using a recognised instrument.

Table 5. *MMAT quality appraisal of included studies*

Study	MMAT category	Criteria met (of 5)	Note
Kadir et al. (2024)	Nonrandomized	4	Strongest; only baseline group equivalence not reported
Baharun et al. (2025)	Quantitative descriptive	5*	*No control group; not assessed by this category's criteria, but discussed separately as a design limitation
Sholiha et al. (2023)	Quantitative descriptive	3	Sample representativeness (single class) not verifiable
Mahendra & Nasir (2025)	Quantitative descriptive	3	Small sample (18 students), not representative
Awaliyah et al. (2026)	Quantitative descriptive (PTK)	3	Small sample (17 students); nonresponse not discussed
Ferdinand &	Nonrandomized	2	Baseline group equivalence not

Yasry (2023)			reported
Nasir & Basit (2021)	Nonrandomized	1	Representativeness and group equivalence unclear
Imron & Fajriyah (2021)	Nonrandomized	1	Control-class outcome data incompletely reported
Cojanu (2025)	Nonrandomized (pilot)	0	Author states the design is inconclusive
Firdaus & Hafidah (2020)	Qualitative	1	Data-analysis process insufficiently reported
Hanifansyah (2025)	Qualitative	1	Retention figures presented without a clear quantitative framework

Synthesis of Findings

Techniques Used

Based on the inductive categorisation described in the Research Methods section, the techniques used in the included studies can be grouped into three broad channels: sound, imagery, and narrative. Four studies used rhymes or songs, two used the keyword mnemonic, and five used story association, visualisation, acronyms, or word grouping (several studies used more than one technique). One study used short mnemonic phrases for numeral rules, and one did not specify its technique. In two studies, mnemonics were combined with other strategies (muscle memory in Hanifansyah, 2025; visualisation and other strategies in (Baharun et al., 2025a), so the effect of mnemonics alone cannot be isolated.

Vocabulary (7 studies)

Five studies reported achievement figures: a posttest mean of 85.59 in the experimental class (Nasir & Basit, 2021), 90.90 (Imron & Fajriyah, 2021), an increase from 41.67 to 77.22 (Mahendra & Nasir, 2025b), from 58.3 to 81.4 (Baharun et al., 2025b), and from 84.94 to 88.41 across two cycles of classroom action research (Nur Iffa Awaliyah et al., 2026). (Ferdinand & Yasry, 2023b) reported a score increase from 49.00 to 86.00, while (Firdaus & Hafidah, 2020) reported qualitative results without score data.

Grammar (4 studies)

(Sholiha et al., 2023b) reported an increase from 61.50 to 81.67, (Kadir et al., 2024b) reported a calculated t of 4.5 above the table t , and Hanifansyah (2025) reported pattern retention of 45% to 82% descriptively. A pilot study by (Cojanu, 2025b) showed only a positive trend.

Retention and Motivation

Retention was measured as a post-treatment score in all quantitative studies. (Baharun et al., 2025c) used a delayed measurement taken one week after the intervention. Motivation was not measured in any study whose text was accessible. This aspect appeared as qualitative reports of

students' enthusiasm (Firdaus & Hafidah, 2020), confidence and motivation (Baharun et al., 2025c), and increased classroom activeness (Nur Iffa Awaliyah et al., 2026)

Design Strength

Only five studies had a control group (one of which was a pilot), and only Kadir et al. (2024) had accessible between-group comparison statistics. Three studies used a single group, one used classroom action research, and two were qualitatively descriptive. Baharun et al. (2025) and Cojanu (2025) noted the limitations of their designs, namely the absence of a control group in the former and the pilot nature of the latter.

The 11 included studies showed a broadly consistent direction of reported findings, but the strength of the evidence was limited. The following discussion addresses the three research questions: the pattern of techniques, the available evidence on vocabulary and grammatical rules, and the strength of that evidence. It then discusses aspects of retention and motivation, the position of this review within the literature, and implications, limitations, and the research agenda.

Patterns of Mnemonic Techniques in Arabic Language Learning

The techniques used revolve around three channels: sound, imagery, and narratives. Rhyme or song was used in four studies (Firdaus & Hafidah, 2020; Imron & Fajriyah, 2021; Sholiha et al., 2023; Awaliyah et al., 2026), keyword mnemonics in two studies (Nasir & Basit, 2021; Ferdinand & Yasry, 2023), and story association, visualisation, acronyms, or word grouping in five studies (Firdaus & Hafidah, 2020; Mahendra & Nasir, 2025; Hanifansyah, 2025; Baharun et al., 2025; Awaliyah et al., 2026). These figures overlap because many studies combine several techniques within a single treatment. In Arabic language learning, mnemonics are more accurately understood as a family of techniques rather than a single method.

The theoretical foundation commonly cited is Dual Coding Theory, which holds that information encoded through both verbal and visual channels is more easily retrieved (Alibasya et al., 2024). The keyword mnemonic applies this principle by linking a foreign word to a similar-sounding, already familiar word, and its combination with retrieval practice has been tested on second-language vocabulary learning (Chiu & Hawkins, 2023). Both references originate from an English-language context; therefore, their applicability to Arabic is an inference of this review rather than a finding of the included studies. The effectiveness of mnemonics on physics formulas (Jusriana et al., 2022) reinforces the assumption that the mechanism operates across fields but does not prove its transfer to Arabic. Rhyme and song, used in four studies, operate through an auditory-verbal channel and are more difficult to situate within the theory. This group of techniques requires a theoretical framework.

The fit between the technique and type of material is apparent in the pattern of use. The keyword mnemonic was used for vocabulary, while rhyme predominated in sharaf patterns (Sholiha et al., 2023) and in the final cycle of the vocabulary classroom action research (Awaliyah et al., 2026). A singing-based strategy combined with flashcard media has similarly

been reported to strengthen vocabulary retention and student confidence among elementary Arabic learners (Rohmawati et al., 2025). This fit has not been tested directly because no study has compared one technique against another. A plausible explanation for this comes from the structure of the language itself. Arabic words are formed by mapping consonantal roots onto vocalic patterns, and a study of 38 Indonesian learners of Arabic found that second-language learners, like native speakers, are sensitive to the presence of the root when processing speech (Aljasser, 2026). If roots and patterns are indeed salient units for learners, a memory aid built on recurring patterns, such as a rhyme applied to a wazn (morphological pattern), is conceptually consistent with the way words are organised. This is an inference of the present review, as Aljasser's study did not test mnemonics.

Evidence on Vocabulary Learning

The seven vocabulary studies reported broadly similar outcomes despite considerable heterogeneity in design, sample size, intervention duration, and outcome measures across studies. In the study with a control class, the pretest scores of both classes were numerically close (66.07 and 62.18); however, the original study did not report a statistical test of baseline equivalence (e.g. an independent-samples t-test), so this numerical similarity alone cannot rule out pre-existing group differences or other confounding factors. The experimental class posttest mean reached 85.59 (Nasir & Basit, 2021). The class taught through songs achieved a mean score of 90.90 (Imron & Fajriyah, 2021), while the one-group design recorded an increase from 41.67 to 77.22, with an N-gain of 0.6288, among 18 students (Mahendra & Nasir, 2025). A cross-country study of 90 santri (Islamic boarding school students) reported a mean increase from 58.3 to 81.4 (Baharun et al., 2025).

These results align with the critique of unstrategic memorisation, namely that vocabulary memorised through repetition does not automatically stick and needs to be connected to the processing of meaning and text (Mulloh, 2025). Mnemonics provide an associative structure for processing. The magnitudes of the increases across studies are not directly comparable, since the tests, educational levels, durations, and numbers of words differed. The moderate N-Gain category in Mahendra and Nasir (2025) indicates that the treatment produced a meaningful improvement but still leaves room for further gains.

Two vocabulary studies require careful reading. The increase in classroom action research from 84.94 to 88.41 occurred between cycles rather than from a pre-treatment baseline, and the technique changed from story association to rhyme (Awaliyah et al., 2026). The decrease in the number of students below the minimum mastery criterion (KKM) from eight to zero out of 17 students is indeed encouraging but cannot be attributed to a single technique. The qualitative study of female senior madrasah students described more enjoyable and enthusiastic learning (Firdaus & Hafidah, 2020), but without score data. Ferdinand and Yasry (2023) reported a 37-point increase (49.00 to 86.00) in their quasi-experiment, although the original report does not specify whether this reflects the experimental class alone or in comparison with the control class.

Evidence on Grammar Learning

The four grammar studies constitute a smaller and methodologically less developed evidence base than the other three. Rhyme mnemonics combined with a deductive approach raised students' scores from 61.50 to 81.67, with a significant paired t-test (Sholiha et al., 2023). A two-group experiment at an Islamic boarding school reported a calculated t of 4.5, exceeding the table t at the 1% and 5% levels (Kadir et al., 2024). A descriptive study reported pattern retention rising from 45% to 82% and comprehension from 40% to 78% (Hanifansyah, 2025), but these figures emerged from a nonexperimental design. A pilot experiment on numeral rules showed only a positive tendency, with the author affirming that the experiment's limitations precluded a definitive conclusion (Cojanu 2025).

The evidence for grammatical rules is inconsistent. Three of the four studies targeted sharaf, that is, the patterns and vocalisation of wazn (morphological templates), and only one targeted nahwu (syntax), and that was a pilot. The assumption that mnemonics assist with grammatical rules because sharaf demands the memorization of many recurring patterns cannot yet be extended to syntactic rules. Hanifansyah (2025) combined mnemonics with muscle-memory-based practice. This suggests that mnemonics help in remembering rules, while their application still requires practice, although the study's design does not allow the respective contributions of the two to be separated.

Strength of Evidence and Design Validity

The distribution of designs shows that only five studies had a comparison group, one of which was a pilot study (Table 3), a pattern corroborated by the MMAT quality appraisal (Table 5). A review of internal validity states that a pretest-posttest design with experimental and control groups controls for most threats to internal validity (Cahit 2015). Accordingly, the three one-group studies (Sholiha et al., 2023; Mahendra & Nasir, 2025; Baharun et al., 2025) can only be read as evidence of improvement, not as evidence that mnemonics caused it. Baharun et al. (2025) had no control group, and the effects of testing practice, maturation, and teacher attention cannot be ruled out. A summary of the conclusions that can and cannot be drawn according to the design is presented in Table 4.

Table 4. What can and cannot be concluded according to study design

Design group	Studies	What can be concluded	What cannot yet be concluded
Experiment with control (5)	Nasir & Basit; Imron & Fajriyah; Ferdinand & Yasry; Kadir et al.; Cojanu (pilot)	Differences in achievement between the treatment and comparison groups; closest to causal inference	Baseline equivalence of groups needs to be checked per study; pilot is inconclusive
One group (3)	Sholiha et al.; Mahendra and Nasir; Baharun et al.	Scores increased after treatment	The increase cannot yet be confirmed as due to mnemonics

Classroom action research (1)	Awaliyah et al.	Gradual improvement within one class across cycles	Technique changed between cycles; no comparison group
Qualitative descriptive (2)	Firdaus & Hafidah; Hanifansyah	Students' experiences and perceptions, possible mechanisms	Treatment effect not tested; not generalizable

Other threats extend beyond the design type. Two studies combined mnemonics with other strategies (Hanifansyah, 2025; Baharun et al., 2025), so the effect of mnemonics alone could not be isolated. The identified sample sizes were small, ranging from 17 to 90 students per study. The longest follow-up was only one week in Baharun et al. (2025). All studies with accessible results (ten studies) reported positive gains or trends. This pattern of exclusively positive results in the published literature raises the possibility of publication bias; however, this could not be formally assessed within the scope of the present narrative review (for example, through funnel plot analysis or statistical tests for small-study effects) and should therefore be treated as an unconfirmed limitation rather than an established finding.

Retention and Affective Aspects

Retention was measured as a post-treatment score in almost all quantitative studies. Only Baharun et al. (2025) used a delayed measurement, which was one week after the intervention. Claims regarding long-term memory, which underlie the theoretical basis of the keyword mnemonic (Alibasya et al., 2024), are not yet supported by evidence from the included Arabic-language studies. The highest retention result in that cross-country study was recorded for the association strategy, followed by mnemonic visualisation and spaced repetition (Baharun et al., 2025); therefore, the role of mnemonics in retention is best read as part of a package of strategies.

Motivation was not measured with an instrument in any study whose text was accessible. This aspect appeared in qualitative reports of enthusiasm (Firdaus & Hafidah, 2020), confidence and motivation (Baharun et al., 2025; Hanifansyah, 2025), and increased student activity (Awaliyah et al., 2026). Motivation to learn Arabic is influenced by internal factors such as interest and self-efficacy, as well as external factors such as the quality of instruction (Widayanti et al., 2025). Technology-based interventions can increase it without affecting the memory process (Fauzi & Anindiati, 2021). Therefore, the relationship between mnemonics and motivation remains hypothetical: students who feel successful at remembering may be more motivated, but this can only be tested with a validated instrument, such as the Arabic-learning motivation instrument developed in international research (Alshammari et al., 2024).

Position of This Review within the Literature

This review complements three existing literature reviews. The literature study on the use of mnemonic methods in Arabic learning is qualitative and descriptive in nature (Elmi & Setiyawan, 2024). A systematic review of speaking skill learning strategies did not focus on mnemonics (Aini et al., 2025). A systematic review of gamification for non-Arabic speakers shows that this type of evidence mapping is feasible for Arabic language learning (Almelhes, 2024b). The distinctive contribution of this review is to sort techniques, subject-matter domains (vocabulary and grammatical rules), and study design strengths into a single map using both a narrative design-based synthesis and a recognised quality appraisal instrument.

This review does not report effect sizes. A narrative synthesis was chosen because the designs and measurement instruments differed across studies, and the quality of the synthesis was constrained by the quality of the available primary studies (Glaros, 2025). Guidance on conducting systematic reviews emphasises that the questions, search strategy, and selection criteria must be stated clearly so that the process can be examined and reproduced (Neilson & Premji, 2024). This review sought to meet that standard, as shown in Table 1 and the selection flow diagram. The position of mnemonics also needs to be kept proportionate to the text. It functions as a complement, not a replacement, for vocabulary work, which must remain connected to text literacy (Mulloh, 2025b).

Implications

The first implication concerns the teachers. The choice of technique should follow the type of material, namely keyword and visual techniques for concrete vocabulary, and rhyme or song for recurring morphological patterns. This pattern was observed across studies and was not the result of a direct comparison between techniques. Teachers who try these techniques are advised to provide a simple comparison, such as a parallel class or a cycle without treatment, so that learning outcomes are not confirmed solely from a rise in scores (Pirdaus, 2024).

The second implication concerns curriculum development and teacher training. However, the available evidence is not yet sufficient to establish mnemonics as a curricular standard. A more appropriate position is as training material and controlled trials, especially since a similar pattern of reported gains appeared across two national contexts (Baharun et al., 2025), while the quality of instruction remains a determinant of student motivation (Widayanti et al., 2025).

The third implication concerns the researchers. Future studies should separate the effect of mnemonics from accompanying strategies, report the baseline equivalence of groups, and include a comparison group. Evidence on syntax (nahwu) and on reading, writing, or speaking skills remains nearly absent, so it is there that the next empirical contributions are most needed (Cojanu, 2025; Hanifansyah, 2025).

Limitations of This Review

The first limitation concerns the search scope. The search was conducted through a general web search engine, DOAJ, and journal portals, without subscription databases such as

Scopus or Web of Science, and covered only Indonesian- and English-language articles, with keywords expanded iteratively rather than being applied as a single fixed query. The second limitation concerns the study process. Screening was carried out by a single reviewer without a second independent reviewer, and for several studies, the data were taken from abstracts, leaving some information incomplete. The third limitation concerns synthesis. The corpus comprises only 11 heterogeneous studies; therefore, meta-analysis is not possible, and the conclusions remain narrative in nature (Page et al., 2021c). Fourth, the possibility of publication bias could not be formally tested and remains an open, unconfirmed concern rather than an established pattern.

Research Agenda

The most pressing research agenda is for controlled-group experiments that compare mnemonic techniques against one another, such as keywords, rhymes, and storytelling, on the same material. Such studies should include a delayed retention test several weeks after treatment and a validated motivational instrument. The evidence needs to be extended to syntax (nahwu) and other language skills, as well as to larger samples and a wider range of educational levels. Once a sufficient number of comparable studies exist, meta-analysis will provide an estimate of effect size that narrative synthesis cannot produce.

CONCLUSION

Because most of the included studies used non-controlled or single-group designs, the findings summarised below should be read as reported positive outcomes across a heterogeneous set of primary studies rather than as established evidence of mnemonic effectiveness. This review synthesises 11 empirical studies from 2020-2026 on mnemonic strategies for learning Arabic vocabulary (seven studies) and grammatical rules (four studies). The most frequently used techniques are rhymes or songs, keyword mnemonics, story association, visualisation, and acronyms, which are often combined within a single treatment. No study has compared one technique with another.

All studies with accessible results (ten studies) reported improvement or a positive trend for both vocabulary and grammatical rules. However, only five studies had a comparison group, one of which was a pilot, while the rest used a one-group, classroom action research, or qualitative descriptive design. Quality appraisal using the MMAT found that only two studies met most core quality criteria. The evidence base for grammatical rules is smaller and methodologically less developed than that for vocabulary, and there is only a single pilot study on syntax. Long-term retention has not been tested, as the longest follow-up was only one week, and motivation has not been measured with an instrument in any of the studies.

For teachers, mnemonics can be tried as a complement to vocabulary and grammar instruction, with a simple comparison to check the results of the same. Future research should compare techniques on the same material with a control group, include a delayed retention test, and measure motivation using a validated instrument. Evidence on syntax, as well as reading, writing, and speaking skills, also needs to be expanded.

Acknowledgments

The authors thank the researchers whose published studies contributed to this systematic literature review. Their work provided the empirical basis for synthesising evidence on mnemonic strategies in Arabic vocabulary and grammar learning.

Author Contribution Statement

Hikmatul Lailaa contributed to the conceptualisation, literature search, study selection, data extraction, and initial drafting of the manuscript. Rasha Kadry Soliman contributed to the critical review, interpretation of findings, and revision of the manuscript. Ammar Zainuddin contributed to the methodological review, synthesis of findings, and final revision of the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding Statement

The authors received no funding from any party for the research and publication of this article.

Declaration of AI Usage

The authors used artificial intelligence tools to assist with language editing and grammar checking during manuscript preparation. All AI-assisted content was reviewed, revised, and verified by the authors, who take full responsibility for the accuracy, originality, and integrity of the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

- Aini, Q., Wahyu Suci Anggrayani, & Fathul Maujud. (2025). Innovative Learning Strategies Towards Improving Students' Arabic Speaking Skills: A Systematic Literature Review. *Sustainable Jurnal Kajian Mutu Pendidikan*, 8(2), 109–120. <https://doi.org/10.32923/xhxgm221>
- Alibasya, A. A., Aditya, S. F., Putri, L. F., & Dwi Farisandy, E. (2024). Long Term Mnemonic: Keefektifan Teknik Keyword Mnemonic untuk Meningkatkan Memori Jangka Panjang Kosakata Bahasa Inggris. *Jurnal Penelitian Dan Pengukuran Psikologi: JPPP*, 13(1), 7–16. <https://doi.org/10.21009/JPPP.131.02>
- Almelhes, S. A. (2024a). Gamification for teaching the Arabic language to non-native speakers: A systematic literature review. *Frontiers in Education*, 9, 1371955. <https://doi.org/10.3389/educ.2024.1371955>
- Alshammari, A., Shaalan, D., & Al-onazi, B. (2024). Instruments measuring motivation to learn Arabic as a second language: Evidence of validity and reliability. *Humanities and Social Sciences Communications*, 11(1), 887. <https://doi.org/10.1057/s41599-024-03381-y>

- Amila Sholiha & Asa Ismia Bunga Aisyahrani. (2023). The Use of Mnemonic Method and Its Implications for Arabic Learning. *Studi Arab*, 14(1), 29–42. <https://doi.org/10.35891/sa.v14i1.3751>
- Asadi, I. A., Vaknin-Nusbaum, V., & Taha, H. (2023). The Role of Morphological Decomposition in Reading Complex Words in Arabic in Elementary School Years. *Journal of Psycholinguistic Research*, 52(6), 2863–2876. <https://doi.org/10.1007/s10936-023-10024-x>
- Baharun, S., Hanifansyah, N., & Atiyah, N. A. (2025a). A Cross-Cultural Psycholinguistic Study on Arabic Vocabulary Retention: Evidence from Indonesia and Malaysia. *Arabiyat : Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban*, 12(2), 200–214. <https://doi.org/10.15408/a.v12i2.46704>
- Chiu, C.-H., & Hawkins, C. F. (2023). The effectiveness of combining the keyword mnemonic with retrieval practice on L2 vocabulary learning in Taiwanese EFL classes. *Journal on English as a Foreign Language*, 13(2), 452–474. <https://doi.org/10.23971/jefl.v13i2.6313>
- Cojanu, E. B. (2025a). MNEMONICS IN TEACHING ARABIC GRAMMAR: A PILOT CASE STUDY ABOUT THE NUMERAL. *Romano-Arabica*, 25(2025), 41–60. https://doi.org/10.62229/roar_xxv/2
- Elmi, B. A. O., & Setiyawan, A. (2024). Pembelajaran Kosakata Bahasa Arab “Al-Mihnah” dengan Teknik Mnemonik (Teori dan Praktik). *EL-IBTIKAR: Jurnal Pendidikan Bahasa Arab*, 13(2), 114. <https://doi.org/10.24235/ibtikar.v13i2.18536>
- Fauzi, Moh. F., & Anindiati, I. (2021). Improving the Motivation of Students in Arabic Language Learning through Learning Management System. *Izdiyar : Journal of Arabic Language Teaching, Linguistics, and Literature*, 4(3), 257–274. <https://doi.org/10.22219/jiz.v4i3.19110>
- Ferdinand, I., & Yasry, F. A. (2023a). Implementation Of The Keyword Mnemonic Method And Its Effectiveness To Improve Arabic Vocabulary Mastery. *Tadris Al-'Arabiyah: Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban*, 2(1), 16–29. <https://doi.org/10.15575/ta.v2i1.22966>
- Firdaus, S., & Hafidah, S. (2020). Mnemonik: Solusi Kreatif untuk Meningkatkan Kemampuan Menghafal Kosa Kata Bahasa Arab Siswi Madrasah Aliyah Nurul Jadid. *PALAPA*, 8(1), 81–96. <https://doi.org/10.36088/palapa.v8i1.700>
- Glaros, A. G. (2025). Statistical significance, clinical importance and effect sizes: Enhancing understanding of a study's results. *Journal of Oral Rehabilitation*, 52(8), 1185–1192. <https://doi.org/10.1111/joor.13759>
- Hong, Q. N., Gonzalez-Reyes, A., & Pluye, P. (2018). Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the MIXED METHODS APPRAISAL TOOL (MMAT). *Journal of Evaluation in Clinical Practice*, 24(3), 459–467. <https://doi.org/10.1111/jep.12884>
- Imron, A., & Fajriyah, D. F. (2021). Penggunaan Metode Bernyanyi dalam Menghafal Mufrodlat (Kosakata) Bahasa Arab di MI. *Dawuh Guru: Jurnal Pendidikan MI/SD*, 1(1), 41–56. <https://doi.org/10.35878/guru.v1i1.255>
- Jusriana, A., & Jurdah Budiman, A. (2022). THE EFFECTIVENESS OF MNEMONIC LEARNING METHODS ON THE ABILITY OF MEMORYING PHYSICS FORMULA FOR STUDENTS OF CLASS VIII MTS DARUSSALAM. *AL-KHAZINI: JURNAL PENDIDIKAN FISIKA*, 2(2), 127–133. <https://doi.org/10.24252/al-khazini.v2i2.32741>

- Kadir, S. D., Yasin, Z., Abdullah, A., Jundi, M., & Hasibuan, R. (2024a). SHAPING MEMORIES, SHARPENING MINDS: UTILIZING THE MNEMONIC METHOD TO ENHANCE STUDENTS' UNDERSTANDING OF ILM SHARF. *Kitaba*, 2(2), 151–163. <https://doi.org/10.18860/kitaba.v2i2.26434>
- Mahendra, G., & Nasir, A. (2025a). The Effectiveness of an Assistive Memory Strategy (Mnemonic) in Enhancing Arabic Vocabulary Proficiency at the Source of Science Middle School: فعالية استخدام استراتيجيات الذاكرة المساعدة في تحسين إتقان مفردات اللغة العربية في مدرسة منبج العلوم المتوسطة. *EL-MAQALAH: Journal of Arabic Language Teaching and Linguistics*, 6(2), 12–28. <https://doi.org/10.22373/maqalah.v6i2.8249>
- Mulloh, T. (2025a). Integrasi Hafalan Mufrodât Dalam Pembelajaran Bahasa Arab: Upaya Meningkatkan Literasi Teks Arab Siswa Ma Al-Ittihad Malang. *Al-Fikra : Jurnal Ilmiah Keislaman*, 24(1), 104. <https://doi.org/10.24014/af.v24i1.37266>
- Nasir, A., & Basit, M. A. (2021). Ta'lim al-Mufradât al-'Arabiyyah 'alâ Tharîqati al-Istidzkâr Mnemonic lil al-Kalimât al-Miftâhiyyah. *Alibbaa': Jurnal Pendidikan Bahasa Arab*, 2(1), 40–60. <https://doi.org/10.19105/alb.v2i1.3944>
- Neilson, C. J., & Premji, Z. (2024). A study of search strategy availability statements and sharing practices for systematic reviews: Ask and you might receive. *Research Synthesis Methods*, 15(3), 441–449. <https://doi.org/10.1002/jrsm.1696>
- Nur Iffa Awaliyah, Susiawati Susiawati, & Alamsyah Alamsyah. (2026). Penerapan Teknik Mnemonik Untuk Meningkatkan Penguasaan Kosakata Bahasa Arab Siswa Kelas VIII SMPIT Makassar Islamic School. *ENGANG: Jurnal Pendidikan, Bahasa, Sastra, Seni, Dan Budaya*, 6(2), 343–352. <https://doi.org/10.37304/enggang.v6i2.23983>
- 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. (2021a). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. <https://doi.org/10.1136/bmj.n71>
- Pirdaus, P. (2024). Project-Based Learning Model Increases Student Creativity and Learning Outcomes in Pancasila and Citizenship Education Learning. *Journal of Education Research and Evaluation*, 8(1), 173–182. <https://doi.org/10.23887/jere.v8i1.73525>
- Sholiha, A., Rofiqi, M., Muhajir, M., & Sidik, F. F. (2023a). The Effect of Mnemonic Method with a Deductive Approach in Shorof Learning on Students Learning Outcomes. *Arabiyatuna: Jurnal Bahasa Arab*, 7(2 November), 525. <https://doi.org/10.29240/jba.v7i2.7607>
- Widayanti, L., Hajam, H., & Taqiyuddin, T. (2025). Dinamika Motivasi Belajar Siswa dalam Pembelajaran Bahasa Arab dan Strategi Penguatannya. *Aphorisme: Journal of Arabic Language, Literature, and Education*, 7(1), 579–598. <https://doi.org/10.37680/aphorisme.v7i1.9471>