



## Drilling Toward Fluency: Effects of the Audio-Lingual Method on Grade-VII Arabic Speaking at State Islamic Junior High School

Yustika Ayu Rahma<sup>1</sup>, Ade Irma<sup>2</sup>✉

<sup>1</sup>Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, Indonesia

<sup>2</sup>Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, Indonesia

### ABSTRACT

**Purpose** – This study examined how the Audio-Lingual Method (ALM) is implemented and how it affects Arabic speaking skills among seventh-grade learners at State Islamic Junior High School (MTsN) 5 Pandeglang.

**Design/methods/approach** – Using a quasi-experimental nonequivalent control group design with intact classes (experimental  $n=20$ ; control  $n=20$ ), we administered a pretest–treatment–posttest sequence contrasting ALM-based instruction with routine instruction under natural classroom conditions. Speaking performance was assessed with standardized prompts and an analytic checklist, and scores were analyzed with assumption-checked inferential tests and effect sizes.

**Findings** – Results showed a clear posttest advantage for the ALM group over the control (mean difference = 18.55;  $t \approx 4.516$ ;  $p < .001$ ; Cohen's  $d \approx 1.43$ ; Hedges'  $g \approx 1.40$ ;  $\eta^2 \approx .368$ ), alongside a shift toward higher scores (elevated mode, higher floor). These gains suggest that structured pattern drills, rapid corrective feedback, and consistent aural models yield meaningful improvements in accuracy and fluency for heterogeneous junior-secondary learners. Limitations include intact-class assignment, no delayed posttest, and a focus on accuracy/fluency rather than discourse–pragmatic outcomes. Future work should use paired or mixed-effects designs with baseline controls, add delayed and far-transfer measures, test ALM plus brief communicative tasks, and examine dose–response and cost-effectiveness to guide scale-up.

**Research implications** – The large, distribution-shifting effects observed suggest that habit-formation pedagogies such as the Audio-Lingual Method merit renewed theoretical and empirical attention in junior-secondary Arabic. Future studies should test Audio-Lingual Method as a core routine combined with brief communicative tasks, use individual-level models (e.g., mixed effects/ANCOVA), and include delayed and far-transfer outcomes to establish durability and generalizability. Dose–response and cost-effectiveness analyses are further needed to guide scalable implementation in resource-variable schools.

### Article History

Received 25 May 2025

Revised 17 June 2025

Accepted 30 June 2025

### Keyword:

Audio-Lingual Method;  
Arabic speaking skills;  
Quasi-Experimental Study.

## Introduction

Oral proficiency is a pivotal outcome for learners of Arabic in today's interconnected societies, shaping mobility, access to knowledge, and intercultural participation (Keilmuan et al., 2025). Within language education, speaking skills integrate pronunciation, fluency, accuracy, and interactional competence, demanding pedagogies that provide intensive, patterned practice alongside meaningful use (Istiqomah & Sopian, 2025). The Audio-Lingual Method (ALM) remains salient because it systematizes habit formation through repetition, dialogue drills, and



This is an open-access article under the CC BY license

**Corresponding Author:** Ade Irma ([ade.irma@uinbanten.ac.id](mailto:ade.irma@uinbanten.ac.id))

**How to Cite (APA 7<sup>th</sup> Edition):** Yustika, A. R., & Irma, A. (2025). Drilling Toward Fluency: Effects of the Audio-Lingual Method on Grade-VII Arabic Speaking at State Islamic Junior High School. *Lingua: Jurnal Keilmuan dan Kependidikan Bahasa Arab*, 11(1), 20–30. <https://doi.org/10.32678/lingua.v11i1.12102>

immediate feedback that can stabilize segmental and suprasegmental features (Barona-Oñate et al., 2020). At the same time, middle-school contexts present unique developmental and curricular constraints that challenge consistent gains without structured routines (Chen & Zhao, 2022). This study addresses those concerns by examining Audio-Lingual Method usage and its effects on seventh-grade learners at a state Islamic junior high school in Pandeglang, Indonesia.

Prior research in Indonesian secondary settings documents that Audio-Lingual Method can raise fluency and pronunciation while revealing classroom inhibitors and enablers such as time allocation, peer support, learner confidence, teacher energy, and engaging materials. In 2025, Safira and Qodir reported improved fluency and articulation in grade-ten cohorts through a three-stage implementation that also surfaced these contextual factors (Safira & Qodir, 2025). In 2024, Rohadatulaisi and colleagues observed effective outcomes in grade-eight classes when Audio-Lingual Method was organized across planning, enactment, and evaluation, with new-vocabulary load mitigated by joint translation before dialogue practice (Afifah Rohadatulaisi et al., 2024). In the same year, Hasibuan and Zainuddin noted gains in fluency, pronunciation, and confidence at a junior secondary school, although audio-media limitations and variability in pattern imitation persisted (Hasibuan & Zainuddin, 2024). Together, these accounts suggest that structured drill-based routines can deliver perceptible benefits while remaining sensitive to local constraints.

Evidence from experimental and classroom-action designs further strengthens causal claims about Audio-Lingual Method impact on speaking. In 2022, Suryani and collaborators found that a class taught with Audio-Lingual Method outperformed a control condition, with a markedly higher posttest mean, thereby rejecting the null hypothesis (Suryani et al., 2022). In 2024, Magfira and colleagues documented improvements in speaking scores across two action-research cycles after sustained Audio-Lingual Method implementation (Magfira, 2024). That same year, Thahir and co-authors reported a rise from a mean of 52.2 to 80.3—an increase of 28.1 percent—across two cycles, reinforcing Audio-Lingual Method effectiveness. These converging results underscore the method's potential to accelerate measurable progress in pronunciation and fluency indices.

Technology-mediated adaptations of audiolingual principles also demonstrate promise for listening–speaking integration. In 2024, Imamah and colleagues described Sam'iyah Syafawiyah at a university language center, combining YouTube input, repetition, questioning, and generalization to enhance listening responses and articulatory precision while acknowledging contextual limits (Imamah et al., 2024). In 2025, Islam and co-authors reported that an Arabic Listening Speaking application produced statistically significant gains in speaking with a medium N-Gain and fostered interactive, enjoyable learning (Islam et al., 2025). Earlier, Aprianto and colleagues in 2020 showed significant effects on speaking linked to sustained exposure to native-speaker input, aligning with Audio-Lingual Method emphasis on accurate aural models (Aprianto et al., 2020). These studies indicate that digital audio and curated exemplars can reinforce habit formation beyond the physical classroom.

Scholars have also problematized the scope of Audio-Lingual Method, arguing that improvements in sentence-level accuracy and basic pronunciation may not automatically transfer to higher-order communicative performance. In 2024, Qodir and Rido'i observed that while Audio-Lingual Method stabilized grammar and segmental features, it required pairing with contextualized communicative activities and targeted teacher development to advance discourse-level competence (Qodir & Rido'i, 2024). Complementary literature reviews connect bilingual practices—including

translanguaging—with enriched vocabulary growth, syntactic development, and communicative confidence, while calling for curriculum design and teacher training attuned to dual-language realities. In 2025, Zikrillah and collaborators emphasized that bilingual approaches, when aligned with audiolingual supports, can scaffold both form and meaning in adolescent learners (Zikrillah et al., 2025). This conversation positions Audio-Lingual Method as necessary but not sufficient for comprehensive oral development.

Despite encouraging findings, several limitations and gaps remain salient for the junior secondary context addressed here. Many reports aggregate outcomes without disaggregating components such as fluency, segmental accuracy, suprasegmental control, and interactive turn-taking, limiting pedagogical specificity. Constraints frequently noted—vocabulary load, audio-media scarcity, uneven pattern-mimicry skill—are not consistently modeled as mediators or moderators of treatment effects. There is also limited evidence tailored to seventh-grade cohorts in state Islamic schools in Banten Province, where curricular pacing and resource profiles may differ from other sites. Consequently, further study is warranted to clarify implementation protocols and effect magnitudes within this underreported setting.

The present study investigates how the audiolingual method is used and how it influences the speaking skills of seventh-grade students at State Islamic Junior High School (MTsN 5 Pandeglang). Specifically, it examines changes in fluency, pronunciation accuracy, and overall speaking performance while documenting implementation routines and classroom supports. By foregrounding vocabulary-load management, peer-support dynamics, teacher energy, and audio-media availability, the study aims to illuminate practical levers for sustained gains. The anticipated contribution is twofold: first, a classroom-embedded implementation model for junior secondary Arabic that integrates audiolingual routines with supportive scaffolds; second, empirical estimates of impact that inform curriculum design, teacher development, and resource planning in comparable schools. In doing so, the study seeks to refine how habit-formation pedagogies translate into functional speaking growth for early adolescent learners in Islamic secondary education.

## Methods

This study employed a quasi-experimental methodology using a nonequivalent control group design to estimate the causal effect of the Audio-Lingual Method on Grade VII speaking skills under natural classroom conditions. The research was conducted at Madrasah Tsanawiyah Negeri 5 Pandeglang with two intact classes purposively selected: VII “B” as the experimental group ( $n=20$ ) and VII “C” as the control group ( $n=20$ ). Procedures comprised a pretest–treatment–posttest sequence in which the experimental group received Audio-Lingual Method-based instruction centered on pattern drills and dialog practice, while the control group followed routine instruction; classroom observations were conducted throughout to document implementation fidelity and learner behavior.

Two instruments were used: (i) a researcher-developed behavioral checklist capturing three indicators of speaking performance (accuracy/smoothness, fluency, and kinesic features/gestures) and (ii) a dialog text aligned to the “home” topic to standardize practice prompts; observations captured attitudinal responses to Audio-Lingual Method as well. Audio playback equipment and printed task sheets supported delivery, and IBM SPSS Statistics 25.0 was used for data processing, assumption testing, and inferential analyses.

Data collection included performance tests scored at pretest and posttest for both groups and structured observations during instruction; test scores constituted the primary outcome variable for statistical comparison. Instrument validity was addressed through content validation procedures to ensure the checklist measured the intended constructs of speaking performance.

Reliability was evaluated via Cronbach's alpha on pilot data ( $n=20$ ), yielding  $\alpha=0.835$ , which indicates good internal consistency under commonly accepted thresholds (e.g.,  $\geq 0.80$ ). Assumption checks involved tests of normality for score distributions and homogeneity of variances between groups; when assumptions held, independent-samples t-tests were conducted (with Welch's correction applied if variances were unequal), and if normality was violated the Mann-Whitney test was used. To gauge practical significance, a coefficient of determination (R-square) from simple linear regression quantified the proportion of variance in speaking scores attributable to Audio-Lingual Method exposure.

All hypothesis tests used a two-tailed  $\alpha=0.05$ , with the null hypothesis stating no effect of Audio-Lingual Method on speaking skills and the alternative positing a positive effect consistent with the study objective. Standard procedures to enhance rigor included standardized administration protocols, common prompts across groups, rater guidance for checklist scoring, and triangulation of test data with observation field notes.

## Results

This section presents the principal findings on the effect of the audiolingual method on the speaking skills of seventh-grade students at State Islamic Junior High School (MTsN) 5 Pandeglang, based on a sequence of descriptive and inferential analyses. The exposition begins with baseline (pretest) statistics indicating comparable initial abilities between the experimental and control groups, followed by posttest changes that map each group's performance shift. Next, between-group comparisons at posttest confirm a significant advantage for the experimental group, reinforced by a large effect size and a precise confidence interval as evidence of the intervention's practical strength. To enrich interpretation, an intra-group gains summary is provided to reveal the redistribution of performance—not merely an increase in averages—most notably the concentration of high scores within the experimental group. Overall, the results are organized in tables so that the empirical argument is clear, transparent, and traceable from data description to pedagogical implications.

### 3.1. Experimental Group Pretest and Posttest

This section presents the experimental group's pretest–posttest results as the internal basis for comparison within the group. The experimental group's pretest summary is as follows:

**Table 1.** Experimental group pretest descriptive statistics

Measure	Value
Valid N	20
Mean	46,95
Median	47
Modus	33
Std. Deviation	11,260
Varians	126,787
Minimum	33
Maximum	67

This table portrays students' initial speaking ability before the audiolingual treatment. A mean of 46.95 with a median of 47 indicates a distribution centered around the middle, whereas a mode of 33 suggests a cluster of students at the lower end—evidence of a noticeable “left tail” or concentration of low scores. The standard deviation of 11.260 (variance 126.787) reflects heterogeneous starting abilities: inter-student variation is broad, implying the need for differentiated scaffolding. The 33–67 range supports this conclusion: there is no ceiling effect, but a potential floor effect among students starting from low baselines. Pedagogically, this profile creates room for a method emphasizing habit formation (pattern drills, repetition, and modeling) to boost phonological control, fluency, and confidence. The experimental group's posttest summary is presented below:

**Table 2.** Experimental group posttest descriptive statistics

Measure	Value
Valid N	20
Mean	74,65
Median	73
Modus	80
Std. Deviation	10,917
Varians	119,187
Minimum	53
Maximum	93

After the audiolingual intervention, the mean rose to 74.65 and the median to 73, signaling a collective shift into the competent–proficient band. The mode of 80 is highly informative: the most frequent score after treatment sits in the high category, indicating that improvement was not limited to a few students but clustered broadly around strong performance. The standard deviation decreased slightly to 10.917, showing marginally tighter dispersion despite substantial mean growth; in other words, quality improved while maintaining consistency. Notably, the score range shifted to 53–93, indicating that the performance floor lifted (minimum now 53) without signs of a serious ceiling effect (maximum 93). This pattern aligns with the Audio-Lingual Method: structured repetition and pattern practice consolidate articulatory accuracy and fluency, yielding substantial and evenly distributed gains.

### 3.2. Control Group Pretest and Posttest

This section presents the control group's pretest–posttest results as the comparative benchmark for the experimental group. The control group's pretest summary is as follows:

**Table 3.** Control group pretest descriptive statistics

Measure	Value
Valid N	20
Mean	45,75
Median	43,50
Modus	40
Std. Deviation	17,048
Varians	290,618
Minimum	20
Maximum	73

At baseline, the control group had a mean of 45.75 and a median of 43.50, slightly lower than the experimental group but substantively comparable. A mode of 40 suggests a lower central cluster than the experimental group, implying that more students started in the basic range. The standard deviation of 17.048 (variance 290.618) is much larger than in the experimental group, indicating internal imbalance:

some students were very weak while others were relatively strong at baseline. The 20–73 range underscores this heterogeneity and points to challenges for routine teaching without intensive differentiation. The control group's posttest results are summarized below:

**Table 4.** Control group posttest descriptive statistics

Measure	Value
Valid N	20
Mean	74,65
Median	73
Modus	80
Std. Deviation	10,917
Variances	119,187
Minimum	53
Maximum	93

Following routine instruction (without the audiolingual method), the mean increased to 56.10 and the median to 60, indicating expected improvement attributable to regular curriculum exposure and repeated practice. The mode remained 40, signaling that a subset of students persisted at a low level; this indicates uneven improvement—gains did not lift all students uniformly. The standard deviation decreased from 17.048 to 14.772, meaning dispersion narrowed slightly but remained larger than in the experimental group; the control group's final performance was thus more heterogeneous. The 40–87 range shows progress (the floor rose from 20 to 40) but not nearly as sharply as in the experimental group. In practical terms, routine instruction elevates outcomes, yet it is less effective at shifting the “mass of performance” into high-score categories compared with audiolingual instruction.

### 3.3. Summary Comparison: Experimental vs. Control

The pretest–posttest comparison for both groups is presented below:

**Table 5.** Summary of mean pretest, mean posttest, and mean gain

Group	Mean Pretest	Mean Posttest	Mean Gain
Experimental	46,95	74,65	<b>+27,70</b>
Control	45,75	56,10	<b>+10,35</b>

This comparative table underscores the treatment effect: the experimental group increased by +27.70 points (46.95 → 74.65), whereas the control group rose by only +10.35 (45.75 → 56.10). The differential growth indicates that the audiolingual method provides a net benefit beyond the natural improvement observed in the control group. Substantively, the experimental group's relative increase ( $\approx +59\%$ ) is nearly triple that of the control group ( $\approx +23\%$ ), reinforcing the method's strong effectiveness within the same intervention window. Moreover, distributional characteristics (a high mode in the experimental group vs. a persistently low mode in the control group) reveal a more favorable redistribution of performance in the experimental condition: students did not merely raise the mean; they clustered in the higher score range. Thus, the table visualizes not only “higher” outcomes but also a more uniform shift toward strong performance for the group receiving the Audio-Lingual Method.

### 3.4. Treatment Effectiveness (Posttest Welch's t-test and Effect Sizes)

**Table 6.** Summary of mean pretest, mean posttest, and mean gain

Group	Mean	Std. Deviation	N
Experimental	74.65	10.917	<b>20</b>
Control	56.10	14.772	<b>20</b>

At posttest, the mean difference of 18.55 yielded  $t \approx 4.516$ ;  $p \approx 6.85 \times 10^{-5}$  ( $p < .001$ ) with a 95% CI of [10.21, 26.89], indicating a highly significant advantage for the

experimental group with good estimation precision. Cohen's  $d \approx 1.43$  (Hedges'  $g \approx 1.40$ ) indicates a practically large effect—not merely statistical significance. Glass's  $\Delta \approx 1.26$  (using the control SD) is consistent: even with greater variability in the control group, the effect remains very strong. The proportion of variance attributable to treatment ( $\eta^2 \approx 0.368$ ) suggests that roughly 37% of the variation in posttest scores can be explained by the instructional approach—substantial for an intervention in regular classroom conditions. Pedagogically, this implies that the Audio-Lingual Method effectively drives the formation of speech habits (pattern practice, drilling, and mimicry) with tangible impacts on fluency and accuracy.

Taken together, the tables display a consistent pattern: (1) balanced baselines, (2) substantial performance gains in the experimental group, (3) robust statistical significance with large practical effects, and (4) a favorable distributional shift (higher mode, raised floor) in the experimental group. This pattern supports the claim that the audiolingual method is effective for seventh-grade State Islamic Junior High School (MTsN) contexts with heterogeneous initial profiles, particularly when the primary goal is the formation of articulatory habits and fluency. For future strengthening, individual-level analyses (paired  $t$  tests,  $d_p$ , ANCOVA) would yield more precise estimates, including tighter control of per-student baseline variation. Practically, teachers can maintain structured pattern drills, enrich phonetic feedback, and add contextual communicative activities to enhance transfer to spontaneous communication. These findings are relevant for school decision-makers considering time allocation, teacher training, and audio media as pillars for sustaining the Audio-Lingual Method.

## Discussion

The present study examined whether and how the audiolingual method (ALM) shapes seventh-grade students' Arabic speaking skills at a state Islamic junior high school in Pandeglang. Framed by research on habit formation, pattern practice, and the role of accurate aural models, the inquiry revisited classic claims that intensive drill can stabilize pronunciation and fluency in early learners (Afifah Rohadatulai et al., 2024; Hasibuan & Zainuddin, 2024; Safira & Qodir, 2025; Zikrillah et al., 2025). We positioned ALM against routine instruction to isolate the method's added value under ordinary school constraints. Prior work in Indonesian secondary settings has reported gains in fluency and articulatory accuracy alongside context-specific enablers and inhibitors, offering a comparative backdrop for interpreting our results. In this light, the research question targeted both measurable outcomes and implementation-level features that might explain differential progress.

The findings showed balanced baselines between groups and a pronounced posttest advantage for the ALM class, with a mean difference of 18.55 points and a large effect size (Cohen's  $d \approx 1.43$ ; Hedges'  $g \approx 1.40$ ). Confidence intervals were relatively narrow, supporting precision in the estimated treatment effect and limiting the likelihood that results reflected sampling noise. Distributional diagnostics were also informative: the experimental group's mode shifted upward and the performance floor rose, indicating not only higher central tendency but also a favorable redistribution of outcomes. By contrast, the control group's improvements were modest and uneven, with a persistent cluster at lower scores. Together, these results suggest that ALM produced substantial, practically meaningful gains in a heterogeneous cohort.

These outcomes cohere with classroom action research and quasi-experimental reports documenting ALM-related gains in fluency, pronunciation, and overall speaking performance in Indonesian schools (e.g., increases across two-cycle PTK studies and significant posttest advantages over controls) (Hasibuan &

Zainuddin, 2024). They also align with studies showing that structured repetition with high-quality input can enhance articulatory precision and listening–speaking integration in Arabic programs using digital media and curated exemplars. At the same time, our results are consonant with critiques that ALM alone may be insufficient for higher-order communicative competence unless paired with contextualized activities and teacher development—an issue we address in our pedagogical implications. Notably, the upward shift in the modal band for the ALM group indicates group-level consolidation, which several prior studies implied but did not describe distributionally. In short, the present evidence strengthens earlier conclusions while adding clarity on how performance redistributes within cohorts.

A plausible explanation for the observed gains is that repeated, high-fidelity modeling and immediate feedback accelerated automatization of segmental and suprasegmental patterns, reducing cognitive load during speech planning (Safira & Qodir, 2025). As basic articulatory routines became procedural, learners could allocate more attention to fluency and turn management, which aligns with the improved central tendency and tightened dispersion. The movement of the minimum score upward suggests that learners with weaker baselines particularly benefited from structured drills and choral practice. Moreover, routine exposure to fixed-response patterns may have built confidence, decreasing disfluencies associated with retrieval and monitoring. These mechanisms are consistent with skill-acquisition accounts of language learning in novice speakers (Suryani et al., 2022).

A second explanatory factor involves the management of vocabulary load and audio-media supports. Where preteaching of key lexis and access to native-speaker models were available, learners likely experienced lower extraneous load and clearer articulatory targets (Magfira, 2024). The modest reduction in the experimental group's standard deviation, despite large mean gains, suggests that such supports helped weaker learners “catch up” without capping stronger learners—a desirable equity effect. In contrast, constraints on audio resources in the control condition may have limited precise feedback on phonetic targets, contributing to the persistent low-mode cluster. These dynamics underscore that ALM's effectiveness is partly contingent on materials quality and delivery fidelity.

Nonetheless, the findings warrant cautious interpretation. First, while the Welch test addressed variance heterogeneity, the study design compared intact classes, leaving open threats from unmeasured classroom-level factors (e.g., teacher enthusiasm, peer norms). Second, without delayed posttests, durability of gains—especially transfer to spontaneous discourse—remains to be established. Third, the assessment focused on accuracy and fluency indices; pragmatic competence and interactional strategies were not directly measured, which narrows the claims that can be made (Aprianto et al., 2020). Finally, although descriptive redistribution patterns are compelling, individual-level modeling (e.g., ANCOVA, mixed-effects) would yield sharper estimates by adjusting for baseline variability. These limits suggest avenues for replication and methodological refinement.

The implications are both pedagogical and programmatic. For classroom practice, ALM can be retained as a core routine for early-stage habit formation, provided it is coupled with short communicative tasks to promote transfer beyond sentence-level control. Teachers should preteach critical vocabulary, ensure sustained access to high-quality audio models, and provide rapid phonetic feedback to maintain fidelity. At the program level, scheduling sufficient drill time, investing in audio infrastructure, and supporting teacher training in error correction and prosody can magnify effects. For research and evaluation, future work should incorporate

paired designs or mixed-effects models, delayed posttests, and discourse-level measures to track durability and breadth of impact. Collectively, these steps can help translate robust gains in accuracy and fluency into stronger spontaneous communication in junior-secondary Arabic.

## Conclusion

This study set out to determine how the audiolingual method is used and how it affects the speaking skills of seventh-grade students at State Islamic Junior High School (MTsN) 5 Pandeglang. Across descriptive and inferential analyses, the audiolingual class substantially outperformed routine instruction, with a posttest mean advantage of 18.55 points, large standardized effects (Cohen's  $d^* \approx 1.43$ ; Hedges'  $g^* \approx 1.40$ ), and an estimated  $\eta^2 \approx 0.368$  indicating meaningful explanatory power. The results further showed a favorable redistribution of performance in the experimental group—an elevated mode and a higher performance floor—signaling not only gains in central tendency but also broader cohort uplift. Pedagogically, these outcomes support the use of structured pattern practice, rapid corrective feedback, and consistent exposure to accurate aural models to consolidate articulatory accuracy and fluency in early adolescent learners. Programmatically, they justify allocating protected drill time, investing in reliable audio infrastructure, and prioritizing teacher training in phonetic feedback and prosody to sustain and scale effects. Several limitations temper these conclusions. The use of intact classes introduces potential confounds from unmeasured class-level factors, and the absence of delayed posttests precludes claims about durability and transfer to spontaneous discourse. The assessment emphasized accuracy and fluency indices, leaving pragmatic, interactional, and discourse-level competencies underexamined. Moreover, while variance-robust testing was applied, individual-level modeling could more precisely adjust for baseline heterogeneity. These constraints point to the need for cautious generalization beyond comparable junior-secondary Arabic contexts. Future research should employ paired or mixed-effects designs with baseline covariate adjustment, include delayed and far-transfer measures, and incorporate discourse-analytic outcomes to evaluate communicative breadth. Comparative trials that blend audiolingual routines with brief, contextualized communicative tasks could test pathways for transferring habit-based gains to spontaneous interaction. Dose–response studies can determine optimal drill intensity and feedback frequency, while implementation research can specify the audio quality, vocabulary preteaching, and teacher coaching required for fidelity. Cost-effectiveness analyses would inform resource allocation for audio media and professional development at the school or district level. Taken together, these steps can translate robust improvements in accuracy and fluency into durable communicative competence for junior-secondary learners of Arabic.

## Declarations

### *Author contribution statement*

Yustika Ayu Rahma conceived and designed the study, coordinated data collection, conducted the primary analyses, and drafted the initial manuscript. Ade Irma refined the study design and instrumentation, ensured methodological rigor and validation, performed supplementary analyses and visualization, and provided critical revisions and overall supervision. Both authors read and approved the final version of the manuscript.

### **Funding statement**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### **Data availability statement**

The datasets generated and/or analyzed during the current study, together with the analysis notes, are available from the corresponding author on reasonable request.

### **Declaration of interests statement**


The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

### **Additional information**

Correspondence and requests for materials should be addressed to [ade.irma@uinbanten.ac.id](mailto:ade.irma@uinbanten.ac.id).

### **ORCID**

Yustika Ayu Rahma 

Ade Irma 

### **References**

- Afifah Rohadatulaisi, J., Susiawati, I., & Hasyim Fanirin, M. (2024). Penggunaan Metode Audiolingual pada Pembelajaran Maharah al-Kalam di Madrasah Tsanawiyah. *INCARE: International Journal of Educational Resources*, 5(4), 357–367. <https://doi.org/10.59689/incare.v5i4.1057>
- Aprianto, A., Ritonga, M., Marlius, Y., & Nusyur, R. (2020). The Influence of Using Audio-lingual Method on Students' Speaking Skill in Madrasah Diniyah Takmiliyah Awwaliyyah. *Izdiyar: Journal of Arabic Language Teaching, Linguistics, and Literature*, 3(2), 147–160. <https://doi.org/10.22219/jiz.v3i2.12514>
- Barona-Oñate, R. V., López-Pérez, S. de los A., López López, J. P., & Mocha-Bonilla, J. A. (2020). Use of E-Learning and Audio-Lingual Method for the Development of Listening Comprehension Skills. In J. Nummenmaa, F. Pérez-González, B. Domenech-Lega, J. Vaunat, & F. Oscar Fernández-Peña (Eds.), *Advances and Applications in Computer Science, Electronics and Industrial Engineering* (pp. 83–98). Springer International Publishing.
- Chen, B., & Zhao, C. (2022). More is less: Homeroom teachers' administrative duties and students' achievements in China. *Teaching and Teacher Education*, 119, 103857. <https://doi.org/https://doi.org/10.1016/j.tate.2022.103857>
- Hasibuan, T. Y., & Zainuddin, D. (2024). Learning Arabic Language with Audio Lingual Media to Improve Speaking Ability at SMP IT al Hijrah 2 Medan. *Alsuna: Journal of Arabic and English Language*, 7(2), 308–322. <https://doi.org/10.31538/alsuna.v7i2.6245>
- Imamah, A., Mahsunah, M., Syaputra, P., Aulia, M., & Fitriani, L. (2024). Implementasi Metode Sam'iyah Syafawiyah pada Maharah Istima' dan Maharah Kalam di PKPBA UIN Malang. *BONJOLE: Borneo Journal of Language and Education*, 4(1), 42–58.
- Islam, M. H., Hasyim, M., & Taubah, M. (2025). The Effectiveness of Arabic Listening

- Speaking Application in Improving Students' Arabic Speaking Skills: A Case Study at MA Al-Hidayah Pasuruan. *Loghat Arabi: Jural Bahasa Arab Dan Pendidikan Bahasa Arab*, 6(1), 291–305. <https://doi.org/10.36915/la.v6i1.436>
- Istiqomah, S. N., & Sopian, A. (2025). Deep Learning Approach for Arabic Vocabulary Mastery in the Digital Era. *Journal of Arabic Language, Literature, and Education*, 6(1), 97–115. <https://doi.org/10.37680/aphorisme.v6i1.7183>
- Keilmuan, J., Arab, B., Zharifah, A., Amrulloh, M. A., & Wahidah, Y. L. (2025). Exploring the Impact of the Show and Tell Method on Student Engagement in Maharah Kalam. *Al-Ittijah*, 17(1), 19–36. <https://doi.org/10.32678/alittijah.v17i1.11197>
- Magfira, A. (2024). Implementasi Metode Audiolingual Terhadap Peningkatan Keterampilan Berbicara Bahasa Arab Di Madrasah Tsanawiyah. *PINISI/ JOURNAL OF ART, HUMANITY & SOCIAL STUDIES*, 4(4), 175–183.
- Qodir, A., & Rido'i, M. (2024). The Audio-Lingual Method as a Teaching Strategy for Arabic. *JoLAL: Journal of Linguistic and Applied Linguistic*, 2(3), 57–63. [digitalpress.gaes-edu.com/index.php/jolal/article/view/334](https://digitalpress.gaes-edu.com/index.php/jolal/article/view/334)
- Safira, B. S., & Qodir, A. (2025). Implementasi Metode Audiolingual Untuk Mengembangkan Maharah Kalam. *JURNAL ILMIAH PENELITIAN MAHASISWA*, 3(4), 465–475. <https://doi.org/10.61722/jipm.v3i4.1129>
- Suryani, R. M., Amir, F. R., & Balgis, L. F. (2022). Efektivitas Metode Audiolingual Dalam Peningkatan Maharah Al-Kalam Bahasa Arab. *Tatsqifiy: Jurnal Pendidikan Bahasa Arab*, 3(1 SE-Articles), 47–56. <https://doi.org/10.30997/tjpba.v3i1.4964>
- Zikrillah, Erlina, E., Rafli, Z., & Amrulloh, M. A. (2025). The Contribution of Bilingualism to the Enhancement of Arabic Listening and Speaking Skills in Language Instruction. *Jurnal Pendidikan Bahasa*, 14(1), 183–200. <https://doi.org/10.31571/bahasa.v14i1.9058>