Daengku: Journal of Humanities and Social Sciences Innovation *ISSN: 2775-6165 (online)*
Vol. 6 No. 3 (2026)
https://doi.org/10.35877/454RI.daengku5004

Adaptive Game-Based Learning and Arabic Vocabulary Achievement: A Quasi-Experimental Evaluation of AI-Supported Instruction

Mochamad Hasyim[a*], S Syarifuddin[a], Lukman Hakim[b], & Mochammad Faiz Sholichin[c]

[a]*Department of Arabic Education, Faculty of Islamic Studies, Universitas Yudharta Pasuruan, Pasuruan, Indonesia*
[b]*Department of Informatics Engineering, Faculty of Engineering, Universitas Yudharta Pasuruan, Pasuruan, Indonesia*
[c]*Universitas KH Abdul Chalim Mojokerto, Mojokerto, Indonesia*

Abstract

This study evaluates the effect of a rule-based AI-supported adaptive game-based learning system on Arabic vocabulary learning among secondary-level students in Indonesia. A quantitative quasi-experimental pretest-posttest control group design was employed with 84 students, consisting of 42 students in the experimental group and 42 students in the control group. The experimental group learned Arabic vocabulary through an adaptive game-based system that adjusted task difficulty based on response accuracy, error patterns, and learning progress, while the control group received conventional vocabulary instruction. Data were analyzed using descriptive statistics, paired-sample t-tests, Welch's t-test for gain-score comparison, effect size estimation, normalized gain analysis, and ANCOVA. The results showed that both groups improved, but the experimental group achieved stronger vocabulary gains than the control group. The mean gain score was 7.07 in the experimental group and 3.64 in the control group. Welch's t-test confirmed a significant between-group difference in gain scores, and ANCOVA showed that the experimental group retained a significant adjusted advantage after controlling for pretest scores. These findings suggest that adaptive game-based learning can support Arabic vocabulary development, although the low N-Gain values and baseline differences require cautious interpretation.

Keywords: AI-supported learning, Adaptive Learning, Game-Based Learning, Arabic Vocabulary, Quasi-Experimental Design

Received: 17 March 2026 Revised: 30 May 2026 Published: 30 June 2026

1. Introduction

The rapid advancement of artificial intelligence in education has encouraged the development of learning systems that can personalize tasks, feedback, and instructional pathways based on learner performance. Recent reviews show that AI-supported learning has moved from broad technological experimentation toward learner modeling, adaptive feedback, and data-informed instructional decision-making (L. Chen et al., 2020; Hwang et al., 2020). In classroom practice, this shift matters because learners rarely enter the classroom with equal prior knowledge, similar learning speed, or identical responses to feedback. AI-supported environments can help teachers respond to these differences by collecting performance information and converting it into instructional decisions. These decisions may include changing the level of difficulty, modifying the sequence of tasks, providing corrective feedback, or offering additional practice. In language education, personalization is especially important because vocabulary learning requires repeated exposure, meaningful use, and gradual consolidation.

Adaptive learning is pedagogically valuable when it is connected to clear learning goals and transparent instructional decisions. Systematic reviews of adaptive learning indicate that AI-enabled systems can support personalized learning pathways, but their value depends on how learner data are translated into meaningful feedback and practice opportunities (Gligorea et al., 2023; Kabudi et al., 2021). Students who already master basic vocabulary need more challenging tasks, while students who still struggle with recognition and recall require more guided repetition. Adaptive systems can move beyond uniform instruction by providing differentiated learning experiences that respond to students'

* Corresponding author.
E-mail address: hasyim@yudharta.ac.id

actual performance. This is particularly relevant in vocabulary learning because a fixed sequence of exercises may be too easy for some students and too difficult for others.

Vocabulary learning occupies a central position in language education because it shapes learners' ability to comprehend input, produce meaningful utterances, and participate in communicative tasks. Vocabulary scholars emphasize that lexical knowledge involves more than knowing word meanings; it also includes form, use, collocation, and contextual application (Nation, 2013; Schmitt & Schmitt, 2020). In Arabic learning contexts, vocabulary mastery is especially important because students must recognize word forms, understand meanings, connect vocabulary with grammatical structures, and use words in appropriate contexts. However, vocabulary instruction in many secondary schools still relies heavily on teacher explanation, memorization, repetition, and written exercises. These practices remain useful, but they often provide limited opportunities for active interaction, immediate feedback, and personalized progression.

Recent vocabulary studies also show that technology can support lexical learning when it strengthens exposure, retrieval, and contextualized practice. Technology-assisted vocabulary learning has been associated with positive learning outcomes, but the effectiveness of digital tools varies according to task quality, feedback, and instructional integration (Alshumrani, 2024; Hao et al., 2021). In many Arabic classrooms, students may remember words for short periods but fail to apply them in context. They may also lose attention when the same instructional pace is imposed on students with different levels of readiness. These conditions create a need for learning systems that support repetition while also sustaining engagement and adapting task difficulty to individual progress.

Game-based learning has been widely discussed as one response to problems of attention and participation in the classroom. Game elements such as points, levels, challenges, feedback, and progression systems can make learning activities more interactive and can encourage students to continue working through tasks (Nadeem et al., 2023; Sailer & Homner, 2020). In vocabulary learning, these elements may create repeated exposure without making practice feel monotonous. Yet the educational value of game-based learning depends on how well game features support learning objectives. A game may increase enthusiasm but still produce limited learning if students focus only on points or speed. For this reason, game-based learning must connect motivation, feedback, repetition, and cognitive processing so that students do not only play but also build stronger vocabulary knowledge.

The effectiveness of game-based learning is strongest when game mechanics are aligned with cognitive goals. Meta-analytic and review evidence indicates that digital games can support learning, but outcomes depend on instructional design, feedback quality, cognitive demand, and the connection between game tasks and learning objectives (Boyle et al., 2016; Clark et al., 2016). More recent studies in digital game-based vocabulary learning also emphasize that learners need meaningful engagement with word meaning, retrieval, and use rather than mere interaction with game features (Zhang et al., 2023; Zou et al., 2021). This evidence suggests that game-based vocabulary learning should be designed as structured practice, not only as motivational entertainment.

Adaptive game-based learning offers a stronger instructional direction because it combines motivational design with performance-sensitive support. In this model, game features sustain participation, while adaptive mechanisms regulate the level of challenge and feedback. Studies on adaptive and AI-assisted game-based learning show that adaptation can support learning outcomes when task difficulty, feedback, and learner behavior are monitored during gameplay (Chen & Chang, 2024; Hwang & Zhang, 2024). This approach aligns with the view that effective learning occurs when students work on tasks that are neither too easy nor too difficult. If a task is too easy, students may become passive. If a task is too difficult, students may experience overload and disengagement. Adaptive systems can reduce this problem by changing task difficulty in response to learner performance.

The present study uses the term AI-supported adaptive game-based learning in a careful and specific sense. The system examined in this study applied a rule-based adaptive mechanism inspired by reinforcement learning principles, particularly the state-action-reward logic associated with Q-learning (Jang et al., 2019). Learner performance indicators, such as response accuracy and error patterns, functioned as signals for determining the next instructional action. Task difficulty represented the action selected by the system, while correct and incorrect responses functioned as feedback signals. However, the system did not implement a fully optimized Q-learning algorithm with autonomous policy updating. It is therefore more accurate to describe the intervention as a rule-based adaptive learning environment inspired by reinforcement learning principles rather than a full reinforcement learning agent.

Although research on AI in education, adaptive learning, and digital game-based learning has expanded rapidly, empirical studies that evaluate integrated adaptive game-based systems through classroom-based comparative designs remain limited. Reviews of AI in education continue to call for stronger empirical designs, clearer pedagogical grounding, and more attention to ethical and classroom implementation issues (Bond et al., 2024; Garzón et al., 2025).

Many studies emphasize system development, user acceptance, motivational response, or technological affordances. Fewer studies test whether adaptive game-based learning produces stronger learning gains than conventional instruction while controlling for initial ability. This gap is important because technology-enhanced learning should not be evaluated only through novelty or engagement claims. It must also be supported by measurable learning outcomes, appropriate comparison groups, and transparent statistical analysis.

The gap is more visible in Arabic vocabulary learning at the secondary education level. Much of the available research on technology-assisted vocabulary learning focuses on English as a foreign or second language (Hao et al., 2021; Lin & Lin, 2019). Learning Arabic presents different instructional challenges because students often face unfamiliar orthography, morphology, pronunciation patterns, and semantic relationships. In Indonesian Islamic school contexts, Arabic is also connected with religious texts and formal curricula, which makes vocabulary mastery academically and culturally important. However, empirical evaluation of adaptive game-based learning for Arabic vocabulary remains scarce. This study addresses that gap by examining a classroom-based intervention in which an experimental group learned through an AI-supported adaptive game-based system and a control group received conventional vocabulary instruction.

This study contributes to the literature in three main ways. First, it provides empirical evidence on adaptive game-based vocabulary learning in an Arabic language context, which remains underrepresented in current AI-supported language learning research. Second, it applies a pretest-posttest control group design and reports gain-score comparison, effect size, normalized gain, and ANCOVA adjustment to account for baseline differences. Third, it offers a cautious conceptual framing of rule-based adaptivity inspired by reinforcement learning principles. The research questions are: (i) To what extent does AI-supported adaptive game-based learning improve Arabic vocabulary learning outcomes compared with conventional instruction? (ii) Does the experimental group demonstrate greater learning improvement after pretest differences are considered?

2. Literature Review

2.1 AI-supported adaptive learning

AI-supported learning refers to instructional environments that use computational processes to support learner modeling, personalization, feedback, or decision-making during learning. Reviews of AI in education show that adaptive learning, intelligent tutoring, learning analytics, and automated feedback have become central themes in the field (Chen et al., 2022; Tapalova et al., 2022). These systems are pedagogically important because they can respond to learner differences that are difficult to address through uniform classroom instruction. Adaptive learning systems use information about learner performance to modify content sequence, task difficulty, feedback, or practice opportunities. In language learning, this capacity matters because students differ in vocabulary size, processing speed, memory strength, and ability to apply words in context.

Adaptive learning should not be understood only as a technical feature. Its value depends on the instructional logic that connects data, feedback, and learning goals. A system may collect many learner data points but still have limited pedagogical value if the adaptation does not support meaningful practice. Recent AI education reviews emphasize the need for transparent, ethical, and learning-oriented design rather than technology-centered implementation (Holmes et al., 2022; Xaveria et al., 2025). In this study, adaptivity is treated as a pedagogical mechanism that supports differentiated vocabulary practice. The rule-based procedure does not claim full algorithmic autonomy. It uses learner response patterns to adjust task difficulty and to provide additional practice when students need support.

AI-supported adaptivity is also connected with the broader development of learning analytics and teacher-facing decision support. Classroom discourse and learning analytics research indicate that AI can help identify learning patterns, but teachers still need interpretable information to make instructional decisions (Bond et al., 2024; D. Wang et al., 2024). This perspective is important for the present study because the system is not positioned as a substitute for teachers. Rather, it supports instructional differentiation by providing adaptive task sequencing and immediate feedback during vocabulary practice. The teacher remains central in explaining meaning, contextualizing vocabulary, and connecting digital practice with classroom communication.

2.2 Game-based learning and vocabulary development

Game-based learning draws on design elements such as goals, rules, challenge, feedback, progression, rewards, and interaction to support learning engagement. In vocabulary learning, game-based activities can provide repeated exposure to words while reducing the monotony often associated with memorization (Plass et al., 2015; Sailer &

Homner, 2020). This is useful because vocabulary development requires frequent encounters with words across different situations and forms. Game features can also increase task persistence by making practice more immediate and interactive. However, the learning effect of game-based learning depends on whether game mechanics direct students toward meaningful language processing rather than only competition or point accumulation.

Recent studies suggest that game-based learning can enhance motivation and participation, but its educational effect is shaped by instructional design quality. Zhang et al. (Zhang et al., 2023) and Zou et al. (Zou et al., 2021) show that digital game-based vocabulary learning becomes more effective when learner engagement is connected with meaningful task design, feedback quality, and vocabulary processing. Game-based vocabulary practice should move beyond points, rewards, or competition by guiding learners to retrieve meanings, distinguish similar words, connect vocabulary with context, and apply words in meaningful language tasks. Such design helps ensure that game-based learning supports cognitive processing rather than merely increasing classroom participation.

The vocabulary-specific literature increasingly supports the need for deeper and more strategic digital game-based language learning. Studies of digital game-based vocabulary learning show that learner engagement, self-regulation, and task design are closely related to vocabulary development (Zhang et al., 2023, 2025). These findings are relevant because Arabic vocabulary learning requires repeated exposure, recognition of form, semantic discrimination, and contextual application. A game environment can support this process only when the game tasks require students to process vocabulary meaning and use, not merely select answers quickly. For this reason, the present study treats game-based interaction as a structured practice environment that combines challenge, feedback, and progression with adaptive task sequencing.

2.3 Reinforcement-learning-informed adaptivity

Reinforcement learning provides a useful conceptual lens for thinking about adaptive instructional decisions because it frames learning as a sequence of states, actions, and rewards. Q-learning is one common reinforcement learning approach in which an agent learns which actions are useful in particular states through feedback over time (Jang et al., 2019). In educational contexts, learner performance can be interpreted as a state signal, instructional task selection can be interpreted as an action, and response accuracy can function as a reward or feedback signal. This logic helps explain how adaptive systems may select the next learning activity based on prior learner responses.

Studies of reinforcement-learning-informed education suggest that adaptive pathways can be useful when implemented with sufficient data, transparent rules, and careful instructional design. Research on Q-learning for vocabulary practice and personalized adaptive learning indicates that performance-based decision rules can support differentiated learning pathways (Dudiak et al., 2023; Sajja et al., 2024). The current intervention uses this logic cautiously. It does not implement a full Q-learning algorithm with autonomous policy optimization. Instead, it uses a transparent rule-based procedure that increases task difficulty after repeated correct responses and decreases task difficulty after repeated incorrect responses.

This cautious framing is important because AI-supported learning can easily be overclaimed. A system that uses rules to adapt difficulty is not the same as an autonomous reinforcement learning agent. However, rule-based adaptivity can still be pedagogically useful when it is transparent, aligned with learning goals, and empirically evaluated. Mobile and adaptive learning research supports the view that transparent adaptive pathways are practical for classroom implementation, especially where teachers need to understand how students move through learning tasks (J. Wang, 2025; Zaini et al., 2025). The system represents a practical form of reinforcement-learning-informed adaptivity rather than a fully autonomous AI agent.

2.4 Arabic Vocabulary learning

Arabic vocabulary learning requires students to develop knowledge of word meaning, form, pronunciation, semantic relationships, and contextual use. For Indonesian secondary-level learners, Arabic is often learned as a foreign language within formal Islamic education. Vocabulary mastery supports reading, classroom communication, and access to religious and academic texts. Studies on vocabulary teaching and technology-assisted vocabulary learning emphasize that vocabulary development requires repeated practice, meaningful context, and feedback-sensitive instruction (Alshumrani, 2024; Asllani & Paçarizi, 2021). This makes Arabic vocabulary learning a relevant context for examining adaptive game-based instruction.

The evaluative gap concerns both context and design. Contextually, Arabic vocabulary learning remains underrepresented in AI-supported adaptive learning research. Methodologically, many technology-based studies focus on learner perception or one-group improvement without sufficient comparison conditions. A pretest-posttest control group design provides stronger evidence because it compares learning gain under two instructional modes. ANCOVA further strengthens interpretation when baseline differences exist. This study positions itself as an evaluation-oriented investigation of adaptive game-based learning in Arabic vocabulary instruction. It examines not only whether students improved, but whether the experimental group improved more than the control group after pretest differences were considered.

3. Methods

3.1 Research Design

This study employed a quantitative quasi-experimental pretest-posttest control group design (Creswell & Creswell, 2017; Rogers & Revesz, 2019). This design was selected because the intervention was implemented in an authentic classroom setting where random assignment at the individual level was not feasible. The study involved two intact classroom groups. The experimental group received AI-supported adaptive game-based vocabulary learning, while the control group received conventional vocabulary instruction. Both groups completed a pretest before the learning activities and a posttest after the intervention period. The design allowed the study to compare learning progress between groups and to examine whether the experimental group achieved stronger vocabulary improvement than the control group.

The control group design was important for strengthening the credibility of the evaluation. A one-group pretest-posttest design can show improvement after an intervention, but it cannot separate the intervention effect from testing effects, maturation, repeated exposure, or ordinary classroom learning. By including a control group, this study could estimate whether the improvement in the experimental group exceeded the improvement obtained through conventional instruction. Because the groups were intact and not randomly assigned, the study remains quasi-experimental. For that reason, the analysis examined baseline differences and used ANCOVA to compare posttest outcomes after controlling for pretest scores.

3.2 Participants and Context

The study was conducted in an Islamic senior high school in Indonesia where Arabic is taught as a compulsory subject. The school context was relevant because Arabic vocabulary learning forms an important part of the curriculum and is closely linked to reading comprehension, classroom communication, and the ability to understand Arabic texts. Prior to the intervention, vocabulary instruction mainly used teacher explanation, memorization, repetition, and written exercises. These practices provided a familiar instructional baseline for comparing the added value of adaptive game-based learning.

The participants consisted of 84 secondary-level students enrolled in Arabic language learning. The experimental group included 42 students, and the control group included 42 students. Participants were selected using purposive sampling because the study used intact classroom groups that were accessible for the intervention. The use of intact groups reflects the practical constraints of classroom-based educational research, where random assignment is often difficult because schools must maintain existing class structures. Although purposive sampling limits generalization, it supports ecological validity because the intervention was implemented in a natural learning setting.

3.3 Intervention and Procedure

The intervention used an AI-supported adaptive game-based learning system designed to support Arabic vocabulary learning. The system integrated vocabulary recognition, contextual use, immediate feedback, scoring, level progression, and challenge-based tasks. The learning content was aligned with the vocabulary material taught during the study period so that both groups received equivalent content. The experimental group interacted with vocabulary tasks through the adaptive game-based system, while the control group learned the same content through conventional instruction. The main distinction between the two groups was the mode of learning, not the vocabulary topic itself.

The adaptive mechanism used a rule-based procedure. Task difficulty increased after three consecutive correct responses and decreased after two consecutive incorrect responses. Correct responses moved students toward more complex tasks, while repeated incorrect responses triggered easier tasks or additional practice. This rule was designed to keep students working at an appropriate level of challenge. The mechanism was inspired by reinforcement learning principles because learner performance functioned as feedback for selecting subsequent instructional actions (Dudiak et al., 2023; Jang et al., 2019). In classroom implementation, this approach offered a transparent approximation of adaptive decision-making. It did not replace the teacher; rather, it supported differentiated task sequencing and immediate feedback during vocabulary practice.

The learning procedure consisted of five stages. First, all students completed a vocabulary pretest to measure initial vocabulary ability. Second, the instructional treatment was implemented in both groups during the same study period. Third, the experimental group used the adaptive game-based system, while the control group received conventional instruction involving teacher explanation, repetition, vocabulary memorization, written exercises, and corrective feedback. Fourth, all students completed a vocabulary posttest after the intervention. Fifth, the data were analyzed to compare within-group improvement and between-group learning gains.

3.4 Instruments

The main instrument was a vocabulary achievement test used as both a pretest and a posttest. The test measured students' ability to recognize, understand, and apply Arabic vocabulary in contextualized learning situations. The items were aligned with the vocabulary content taught during the intervention. Using the same construct across the pretest and posttest allowed the study to examine changes in vocabulary performance over time. The test focused not only on basic word recognition but also on understanding vocabulary meaning and applying vocabulary in simple contextual tasks (Nation, 2013; Schmitt & Schmitt, 2020).

The vocabulary achievement test was reviewed by experts in Arabic language education and educational technology to ensure content validity. Expert review focused on the relevance of the items to the learning objectives, the clarity of item wording, the appropriateness of vocabulary level for secondary-level students, and the alignment between the test and instructional content. Instrument reliability and item quality were examined to support the accuracy and consistency of the measurement. These procedures helped ensure that the test was aligned with the instructional content and appropriate for assessing students' Arabic vocabulary mastery.

The study also paid attention to the interpretation of learning progress. In addition to raw gain scores, normalized gain was used to examine improvement relative to possible score increase. This measure is useful because two students with similar raw gains may have different room for improvement depending on their pretest scores (Bao, 2006; Marx & Cummings, 2007). However, normalized gain must be interpreted carefully, particularly when baseline scores are already moderate and when intervention duration is limited. For this reason, the study reported both raw gain and N-Gain results rather than relying on a single indicator.

3.5 Data Analysis

Data were analyzed using descriptive and inferential statistical procedures. Descriptive statistics summarized pretest scores, posttest scores, gain scores, normalized gain scores, and standard deviations. Shapiro-Wilk tests were used to assess normality because the sample size in each group was below 50 (Field, 2024). Levene tests were used to assess homogeneity of variance. Paired-sample t-tests examined whether each group improved significantly from pretest to posttest. Cohen's dz was used to interpret within-group effect sizes because the paired design involved repeated measurement of the same students.

Between-group learning progress was analyzed through gain-score comparison. Because the gain score did not fully meet the homogeneity assumption, the Welch independent-sample t-test was used for the main gain-score comparison. A Mann-Whitney U test was also used as a robustness check because the gain and N-Gain distributions were not normally distributed. Cohen's d was used to estimate the magnitude of the between-group effect. ANCOVA was then conducted using posttest scores as the dependent variable, group as the independent variable, and pretest scores as the covariate. This procedure was necessary because the two groups differed significantly at baseline.

4. Result and Discussion

The analysis begins with descriptive statistics, continues with within-group improvement, reports between-group gain-score differences, and concludes with ANCOVA adjustment for baseline nonequivalence.

4.1 Descriptive Statistics

Table 1 shows the descriptive statistics for the experimental and control groups. Both groups improved from pretest to posttest. The experimental group increased from 61.33 to 68.40, while the control group increased from 59.31 to 62.95. The experimental group obtained a larger mean gain score than the control group, with a gain of 7.07 compared with

3.64. The difference in mean gain was 3.43 points. These results indicate that vocabulary scores improved in both instructional conditions, but the experimental group showed a stronger improvement pattern.

The normalized gain values also show a difference between groups. The mean N-Gain was 0.186 for the experimental group and 0.095 for the control group. Based on common N-Gain interpretation, both values fall within the low category. This means that the intervention produced measurable improvement, but the improvement did not reach a high normalized gain level. The low N-Gain values are important because they prevent an exaggerated interpretation of the intervention.

A baseline comparison was conducted to examine whether the two groups differed in their initial vocabulary ability. The Welch independent-samples t-test showed that the experimental group had a significantly higher pretest score than the control group, $t(81.69) = 2.360$, $p = .021$. This baseline difference justified the use of ANCOVA to compare posttest scores after controlling for initial vocabulary ability.

Table 1. Descriptive Statistics of Vocabulary Scores

Group	N	Pretest Mean	SD	Posttest Mean	SD	Gain Mean	N-Gain
Experimental	42	61.33	4.04	68.40	6.67	7.07	0.186
Control	42	59.31	3.80	62.95	6.33	3.64	0.095

4.2 Within-Group Improvement

Table 2 presents the paired-sample t-test results for each group. The experimental group showed a significant increase from pretest to posttest, with a mean difference of 7.07 points, $t(41) = 8.757$, $p < .001$. The within-group effect size was large, with Cohen's dz = 1.351. This result indicates that students in the experimental group experienced substantial improvement after participating in the adaptive game-based vocabulary learning activities.

The control group also showed a significant increase from pretest to posttest, with a mean difference of 3.64 points, $t(41) = 6.846$, $p < .001$. The within-group effect size was also large, with Cohen's dz = 1.056. This finding is important because it shows that conventional instruction also produced learning progress. The intervention should not be interpreted as the only source of vocabulary improvement.

Table 2. Paired-sample t-tests for within-group improvement

Group	Pretest Mean	Posttest Mean	Mean Difference	t	df	p-value	Cohen’s dz
Experimental	61.33	68.40	7.07	8.757	41	< .001	1.351
Control	59.31	62.95	3.64	6.846	41	< .001	1.056

4.3 Between-Group Differences

Table 3 reports the between-group comparison of gain scores using the Welch t-test. The experimental group obtained a higher mean gain score than the control group. The gain-score difference was statistically significant, $t(70.96) = 3.545$, $p = .001$, with a medium-to-large effect size, Cohen's d = 0.767. This result indicates that the experimental group achieved greater vocabulary improvement than the control group during the study period.

A Mann-Whitney U test was used as a robustness check and confirmed the significant gain-score difference between groups, $U = 1259.00$, $p = .001$, with a rank-biserial correlation of 0.427. The robustness check supports the Welch result and strengthens confidence that the gain-score difference did not depend only on parametric assumptions. Because the experimental group had a higher baseline score, the gain-score result should still be interpreted together with ANCOVA adjustment.

Table 3. Comparisons of Vocabulary Scores

Comparison	Experimental Mean	Control Mean	Mean Difference	t	df	p-value	Cohen’s d
Gain score	7.07	3.64	3.43	3.545	70.96	.001	0.767

4.4 ANCOVA Adjustment for Baseline Difference

Because the experimental group started with a significantly higher pretest score, ANCOVA was used to compare posttest outcomes after controlling for pretest differences. The adjusted posttest mean was 67.16 for the experimental group and 64.20 for the control group. The group effect remained statistically significant after adjustment, indicating that the experimental group retained an advantage even when initial vocabulary ability was considered.

The classical ANCOVA result showed a significant group effect, $F(1, 81) = 9.020$, $p = .004$, partial eta squared = .100, as shown in Table 4. The partial eta squared value indicates a meaningful but not excessive group effect. This result supports the interpretation that the intervention contributed to stronger posttest performance, but it also reinforces the need for careful interpretation because the groups were not randomly assigned.

Table 4. ANCOVA Posttest Scores

Source	F	df	p-value	Partial η^2
Group effect	9.020	1, 81	.004	.100

4.5 Interpretation of Learning Gains

The findings indicate that AI-supported adaptive game-based learning produced stronger Arabic vocabulary learning progress than conventional instruction. Both groups improved from pretest to posttest, but the experimental group showed a larger gain score, higher posttest performance, and higher N-Gain. This pattern is consistent with recent evidence that digital game-based vocabulary learning can support receptive and productive vocabulary knowledge when learners engage with structured vocabulary tasks (Jia et al., 2024). In the present study, the stronger gains in the experimental group suggest that vocabulary improvement was supported not only by game elements, but also by repeated practice, immediate feedback, and adaptive task progression.

Kazu and Kuvvetli (2023) similarly found that digital game-based language learning can contribute to vocabulary acquisition when instructional tasks are meaningfully connected to learners’ engagement and classroom learning goals. This perspective helps explain why the experimental group outperformed the control group. The adaptive game-based system did not simply present vocabulary items in a more entertaining format; it organized practice through challenge, feedback, scoring, and level progression. These features may have helped students sustain attention during repeated vocabulary practice and respond more actively to learning tasks.

Nevertheless, the baseline difference between groups requires careful interpretation. Because the experimental group began with a higher pretest score, the strongest evidence comes from the gain-score comparison and ANCOVA results rather than from posttest comparison alone. The ANCOVA result strengthens the interpretation that the intervention had instructional value because the experimental group still showed a significant adjusted advantage after pretest scores were controlled. For a classroom-based quasi-experiment, this adjusted comparison provides more credible evidence than a simple posttest comparison because it reduces the risk of attributing all posttest differences to the intervention. These findings support the central argument of the study that adaptive game-based learning can produce stronger vocabulary learning progress when game interaction is combined with performance-sensitive task sequencing and immediate feedback.

The control group also improved significantly, and this finding should not be overlooked. Students in the control group learned the same vocabulary content and received teacher guidance, repetition, written exercises, and corrective feedback. These conventional activities remain pedagogically useful, especially in Arabic classrooms where vocabulary instruction often depends on teacher explanation and guided practice. The intervention should not be interpreted as replacing conventional instruction. Its contribution lies in strengthening the pace, feedback, and differentiation of vocabulary practice.

The low N-Gain values also require cautious interpretation. Although the experimental group significantly outperformed the control group, the normalized gain remained low. Normalized gain is useful because it explains proportional learning improvement rather than merely describing raw score increase (Bao, 2006; Marx & Cummings, 2007). Thus, the intervention can be interpreted as producing stronger progress than conventional instruction, but not as generating a high level of proportional improvement. Future versions of the system should strengthen spaced repetition, feedback depth, contextualized sentence production, and individual learning path analytics.

4.6 Adaptive Mechanism and Instructional Design

The learning advantage in the experimental group can be explained by the combination of adaptive task sequencing and game-based feedback. The rule-based adaptive mechanism increased difficulty after repeated correct responses and reduced difficulty after repeated incorrect responses. This procedure helped align task challenge with learner performance, a principle that is consistent with adaptive learning research on personalized sequencing and feedback

(Gligorea et al., 2023; Kabudi et al., 2021). In classroom practice, this alignment matters because students rarely progress at the same pace. Some students need additional repetition, while others require more challenging tasks to sustain learning.

Chowdhury et al. (2024) emphasize that digital game-based language learning can create meaningful vocabulary learning experiences when learners are given opportunities for agency, contextualized learning, and generative engagement. This insight is relevant to the present study because vocabulary learning was supported through interactive tasks rather than passive memorization. The adaptive system encouraged students to respond to vocabulary items, receive feedback, and move through task levels according to their performance. Such a structure may reduce the mismatch between learner readiness and task difficulty, which is a common limitation in uniform classroom instruction.

The adaptive mechanism used in this study was deliberately framed as rule-based and inspired by reinforcement learning principles. The system used learner performance as feedback for selecting subsequent tasks, reflecting the state-action-reward logic associated with reinforcement learning (Jang et al., 2019). Yet the system did not implement a full Q-learning model with autonomous policy optimization. This distinction is important because AI-supported learning can easily be overclaimed. The intervention should be understood as a transparent, classroom-feasible form of reinforcement-learning-informed adaptivity rather than a fully autonomous AI agent.

The role of immediate feedback also helps explain the learning process observed in the experimental group. In foreign language online assignments, feedback timing can influence how learners process errors and improve performance (Lu et al., 2023). Immediate feedback in the adaptive game-based system may have helped students recognize mistakes and continue practicing without waiting for delayed correction. However, immediate feedback alone is not sufficient. For vocabulary learning, feedback should not only indicate whether an answer is correct, but also guide students toward meaning, usage, and contextual application.

Game-based elements may also have supported the learning process by sustaining attention and encouraging task persistence. Scoring, progression levels, immediate feedback, and challenge-based activities can make vocabulary practice more active and less monotonous when game mechanics are aligned with vocabulary learning tasks (Jia et al., 2024; Kazu & Kuvvetli, 2023). The low N-Gain values show that game features should not be treated as sufficient by themselves. Motivation and participation must be connected with deeper cognitive processing, including retrieval, semantic discrimination, sentence-level use, and repeated exposure over time.

4.7 Pedagogical and Practical Implications

For Arabic language teaching, the findings suggest that adaptive game-based learning can be used as a supplementary strategy for differentiated vocabulary practice. In classrooms with mixed ability levels, a single sequence of vocabulary exercises may not adequately respond to students' different levels of readiness. Adaptive task progression can help address this problem by offering practice that corresponds more closely to students' performance. Nevertheless, teacher guidance remains essential. Teachers are still needed to explain vocabulary meanings, model pronunciation, provide examples, correct errors, and connect digital practice with spoken and written classroom activities.

The integration of AI in language education should also be viewed through pedagogical design rather than technological novelty alone. Zhu and Wang (2025) argue that AI in language education needs to be examined in relation to learner engagement, instructional design, and future implementation challenges. This perspective is important for interpreting the present findings. The adaptive game-based system showed instructional potential, but its value depends on how well it is integrated into classroom teaching and extended into oral, written, and text-based Arabic learning activities.

From an educational technology perspective, adaptive learning systems should make the adaptation process transparent, pedagogically meaningful, and easy for teachers to monitor. Teachers need to understand why students receive certain tasks, how task difficulty changes, and which learners require additional support. For this reason, future systems should include teacher-facing dashboards or log summaries that display accuracy, error patterns, level movement, time on task, and learning progress. Such information would make the system useful not only as a student practice tool but also as a source of instructional feedback for teachers.

Alfredo et al. (2024) emphasize the importance of human-centred learning analytics and AI in education because learner data, automation, and teacher decision-making need to remain aligned with human control, transparency, and educational trust. This argument strengthens the practical implication of the present study. Adaptive systems should not operate as hidden black boxes. Instead, they should provide interpretable information that helps teachers make instructional decisions. In Arabic vocabulary learning, this is especially important because students may recognize a word in a game but still need teacher support to understand its morphology, pronunciation, syntactic function, and contextual meaning.

At the policy level, the study highlights the need to evaluate AI-supported learning through evidence rather than novelty. Schools may be attracted to AI-based tools because they appear innovative, but adoption should depend on whether the technology improves learning, supports teachers, protects student data, and fits classroom conditions. Ethical and transparent implementation is particularly important when learner data are used to guide instructional decisions (Holmes et al., 2022). Policy support should include teacher training, infrastructure readiness, content alignment, data privacy protection, and continuous evaluation of learning outcomes.

5. Conclusion

This study found that AI-supported adaptive game-based learning produced stronger Arabic vocabulary learning progress than conventional instruction in a pretest-posttest control group design. The experimental group obtained a higher mean gain, posttest performance, and N-Gain than the control group. The gain-score comparison confirmed a significant between-group difference, while the ANCOVA result showed that the experimental group retained a significant adjusted advantage after pretest differences were controlled.

The main contribution of this study lies in its evaluation of rule-based adaptive game-based learning in an Arabic vocabulary context. By combining adaptive task sequencing, immediate feedback, and game-based interaction, the intervention offered differentiated vocabulary practice in a classroom setting. The study also shows that technology-supported language learning should be judged not only by engagement or novelty, but also by comparative outcome data, effect size interpretation, and appropriate statistical control.

These findings should be interpreted with caution. The study used intact classroom groups, the experimental group had a higher pretest score, and the N-Gain values remained low. The dataset also focused on vocabulary test outcomes and did not include complete system log-data. Future research should use randomized or matched-group designs, delayed posttests, log-data analysis, and comparisons between adaptive and non-adaptive game-based learning to clarify how game features, feedback, and adaptive sequencing contribute to vocabulary development.

References

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, *6*, 100215. https://doi.org/10.1016/j.caeai.2024.100215

Alshumrani, H. A. (2024). Unveiling vocabulary teaching and learning beliefs of teachers and learners in an EFL context. *Asian-Pacific Journal of Second and Foreign Language Education*, *9*(1), 20. https://doi.org/10.1186/s40862-023-00242-0

Asllani, H., & Paçarizi, R. (2021). Enhancing Second Language Incidental Vocabulary Learning Through Technology. *Journal of Educational and Social Research*, *11*(4). https://doi.org/10.36941/jesr-2021-0081

Bao, L. (2006). Theoretical comparisons of average normalized gain calculations. *American Journal of Physics*, *74*(10), 917–922. https://doi.org/10.1119/1.2213632

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, *21*, 1–41. https://doi.org/10.1186/s41239-023-00436-z

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, *94*, 178–192. https://doi.org/10.1016/j.compedu.2015.11.003

Chen, C.-H., & Chang, C.-L. (2024). Effectiveness of AI-assisted game-based learning on science learning outcomes, intrinsic motivation, cognitive load, and learning behavior. *Education and Information Technologies*, *29*, 18621–18642. https://doi.org/10.1007/s10639-024-12553-x

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, *8*, 75264–75278. https://doi.org/10.1109/access.2020.2988510

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two Decades of Artificial Intelligence in Education: Contributors, Collaborations, Research Topics, Challenges, and Future Directions. *Educational Technology & Society*, *25*(1), 28–47. https://doi.org/10.30191/ETS.202201_25(1).0003

Chowdhury, M., Dixon, L. Q., Kuo, L.-J., Donaldson, J. P., Eslami, Z., Viruru, R., & Luo, W. (2024). Digital game-based language learning for vocabulary development. *Computers and Education Open*, *6*, 100160. https://doi.org/10.1016/j.caeo.2024.100160

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research*, *86*(1), 79–122. https://doi.org/10.3102/0034654315582065

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Dudiak, D., Szabóová, M., & Magyar, J. (2023). Exploring Q-Learning in Social Robots for English-Slovak Vocabulary Learning. *2023 IEEE 23rd International Symposium on Computational Intelligence and Informatics (CINTI)*, 25–30. https://doi.org/10.1109/cinti59972.2023.10382107

Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. SAGE Publications.

Garzón, J., Patiño, E., & Marulanda, C. (2025). Systematic review of artificial intelligence in education: Trends, benefits, and challenges. *Multimodal Technologies and Interaction*, *9*(8), 84. https://doi.org/10.3390/mti9080084

Gligorea, I., Cioca, M., Oancea, R., Gorski, A., Gorski, H., & Tudorache, P. (2023). Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature Review. *Education Sciences*, *13*(12). https://doi.org/10.3390/educsci13121216

Hao, T., Wang, Z., & Ardasheva, Y. (2021). Technology-Assisted Vocabulary Learning for EFL Learners: A Meta-Analysis. *Journal of Research on Educational Effectiveness*, *14*(3), 645–667. https://doi.org/10.1080/19345747.2021.1917028

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, *32*(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Hwang, G.-J., & Zhang, D. (2024). Effects of an adaptive computer agent-based digital game on EFL students' English learning outcomes. *Educational Technology Research and Development*, *72*, 3271–3294. https://doi.org/10.1007/s11423-024-10396-4

Jang, B., Kim, M., Harerimana, G., & Kim, J. W. (2019). Q-Learning Algorithms: A Comprehensive Classification and Applications. *IEEE Access*, *7*, 133653–133667. https://doi.org/10.1109/access.2019.2941229

Jia, W., Zhang, L., Pack, A., Guan, Y., & Zou, B. (2024). Digital game-based learning's effectiveness on EFL learners' receptive and productive vocabulary knowledge. *Language Learning & Technology*, *28*(1), 1–21. https://doi.org/10.64152/10125/73554

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, *2*, 100017. https://doi.org/10.1016/j.caeai.2021.100017

Kazu, İ. Y., & Kuvvetli, M. (2023). A triangulation method on the effectiveness of digital game-based language learning for vocabulary acquisition. *Education and Information Technologies*, *28*, 13541–13567.

https://doi.org/10.1007/s10639-023-11756-y

Lin, J.-J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *32*(8), 878–919. https://doi.org/10.1080/09588221.2018.1541359

Lu, X., Wang, W., Motz, B. A., Ye, W., & Heffernan, N. T. (2023). Immediate text-based feedback timing on foreign language online assignments: How immediate should immediate feedback be? *Computers and Education Open*, *5*, 100148. https://doi.org/10.1016/j.caeo.2023.100148

Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics*, *75*(1), 87–91. https://doi.org/10.1119/1.2372468

Nadeem, M., Oroszlányová, M., & Farag, W. (2023). Effect of Digital Game-Based Learning on Student Engagement and Motivation. *Computers*, *12*(9), 177. https://doi.org/10.3390/computers12090177

Nation, I. S. P. (2013). Learning Vocabulary in Another Language. In *Cambridge Applied Linguistics* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139858656

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of Game-Based Learning. *Educational Psychologist*, *50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

Rogers, J., & Revesz, A. (2019). Experimental and quasi-experimental designs. In *The Routledge handbook of research methods in applied linguistics* (pp. 133–143). Routledge.

Sailer, M., & Homner, L. (2020). The Gamification of Learning: a Meta-analysis. *Educational Psychology Review*, *32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial Intelligence-Enabled Intelligent Assistant for Personalized and Adaptive Learning in Higher Education. *Information*, *15*(10), 596. https://doi.org/10.3390/info15100596

Schmitt, N., & Schmitt, D. (2020). *Vocabulary in Language Teaching* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108569057

Tapalova, O., Zhiyenbayeva, N., & Gura, D. (2022). Artificial Intelligence in Education: AIEd for Personalised Learning Pathways. *Electronic Journal of E-Learning*, *20*(5), 639–653. https://doi.org/10.34190/ejel.20.5.2597

Wang, D., Tao, Y., & Chen, G. (2024). Artificial intelligence in classroom discourse: A systematic review of the past decade. *International Journal of Educational Research*, *123*, 102275. https://doi.org/10.1016/j.ijer.2023.102275

Wang, J. (2025). Intelligent Education Based on Mobile Learning: Transitioning from Traditional Classrooms to Adaptive Learning Environments. *International Journal of Interactive Mobile Technologies*, *19*(11), 51–65. https://doi.org/10.3991/ijim.v19i11.56057

Xaveria, F., Kristianingsih, D., & Maharani, R. (2025). Artificial intelligence in adaptive education: a systematic review of techniques for personalized learning. *Discover Education*, *4*, 458. https://doi.org/10.1007/s44217-025-00908-6

Zaini, N. A., Wook, T. S. M. T., Khalid, M. N. A., & Noah, S. (2025). User Requirements of Adaptive Learning Through Digital Game-Based Learning: User-Centered Design Approach to Enhance the Language Literacy Development. *International Journal of Advanced Computer Science and Applications*, *16*(9). https://doi.org/10.14569/ijacsa.2025.0160919

Zhang, R., Zou, D., & Cheng, G. (2023). Learner engagement in digital game-based vocabulary learning and its effects on EFL vocabulary development. *System*, *119*, 103173. https://doi.org/10.1016/j.system.2023.103173

Zhang, R., Zou, D., & Cheng, G. (2025). Self-regulated digital game-based vocabulary learning: motivation, application of self-regulated learning strategies, EFL vocabulary knowledge development, and their interplay. *Computer Assisted Language Learning*, *38*(8), 1961–2003. https://doi.org/10.1080/09588221.2024.2344555

Zhu, M., & Wang, C. (2025). A systematic review of research on AI in language education: Current status and future implications. *Language Learning & Technology*, *29*(1), 1–29. https://doi.org/10.64152/10125/73606

Zou, D., Huang, Y., & Xie, H. (2021). Digital game-based vocabulary learning: where are we and where are we going? *Computer Assisted Language Learning*, *34*(5), 751–777. https://doi.org/10.1080/09588221.2019.1640745