THE EFFECTS OF AI-SUPPORTED GAME-BASED LEARNING ON ENGLISH VOCABULARY ACQUISITION AMONG GRADE 10 STUDENTS IN HO CHI MINH CITY

Vu Dinh Thanh
Ho Chi Minh City Open University, Vietnam
Email: thanhvd.248t@ou.edu.vn

Abstract: *This study examined a teacher-guided, AI-supported game-based vocabulary intervention using Curipod and its relationship with Grade 10 students' immediate vocabulary achievement and learning perceptions. A quantitative quasi-experimental pre-test–post-test design involved 46 students from two intact classes at Vinh Loc High School in Ho Chi Minh City. The experimental and comparison groups each included 23 students and received 18 lessons over six weeks using Units 1–3 of English 10: Global Success. The experimental group practised through Curipod prompts, matching and contextual sentence tasks, selected AI-generated feedback, and small-team challenges. The comparison group received teacher explanations, textbook exercises, and printed practice. The experimental group's mean score increased from 7.630 to 8.696, while the comparison group's mean changed from 7.413 to 7.304. The difference in mean gains was 1.174 points in favour of the experimental group ($p < .001$ in the reported student-level analysis). Questionnaire responses from the experimental group were generally favourable, particularly for motivation, enjoyment, perceived usefulness, and technology acceptance. The findings suggest that the combined approach was associated with stronger immediate performance on curriculum-based vocabulary tests in these classes. However, the small convenience sample, one intact class per condition, uncertain test-form comparability, positive-only questionnaire items, and absence of delayed assessment limit causal interpretation and claims about long-term retention or communicative transfer. The study supports further controlled classroom investigation rather than claims about the independent effect of AI.*

Keywords: *AI-supported learning; game-based learning; vocabulary acquisition; Curipod; Vietnamese EFL learners*

Nhận bài: 04/10/2026 *Phản biện: 04/11/2026* *Duyệt đăng: 07/11/2026*

I. INTRODUCTION

Vocabulary is a central component of English learning because students need to recognise word forms, understand meanings, and select words appropriately in context. Remembering a translation or definition, however, does not necessarily mean that a learner can retrieve a word independently or use it accurately in a new sentence. Effective vocabulary instruction therefore combines explicit explanation with repeated encounters, contextualised practice, and opportunities to check understanding. Research on vocabulary learning has highlighted the importance of strategic learning and the depth of word knowledge, rather than treating vocabulary as a list of isolated word–meaning pairs (Afzal, 2019; Ghalebi et al., 2020; Zarfsaz & Yeganehpour, 2021). Game-based learning organises practice around challenges, rules, goals, and feedback. Such features can encourage learners to respond and try again, but points or competition do not automatically produce learning. Outcomes depend on alignment among objectives, tasks, feedback, and teacher facilitation; reviews report variation across designs and learner groups (Hartt et al., 2020; Karakoc et al., 2022; Plass et al., 2015). AI-supported platforms may help teachers prepare prompts, generate examples, and provide feedback. However, generated content can be inaccurate or unsuitable and requires teacher review. This study treats AI as one component of a teacher-guided arrangement, not as an autonomous teacher or a technology separable from task design and classroom interaction. The study was conducted at Vinh Loc High School in Ho Chi Minh City and focused on Grade 10 vocabulary from Units 1–3 of English 10: Global Success. Curipod organised interactive prompts, sentence activities, selected AI-generated feedback, and team challenges. The comparison class studied the same curricular content through teacher explanations, textbook exercises, and printed practice. The study asked: (1) To what extent did AI-supported game-based instruction through Curipod influence immediate vocabulary achievement compared with conventional instruction? (2) What attitudes did experimental-group students report toward the learning activities? Achievement was measured through curriculum-based tests of recognition and controlled contextual completion, not through comprehensive measures of communicative vocabulary use.

II. LITERATURE REVIEW

Vocabulary knowledge connects word form, meaning, grammatical behaviour, and contextual appropriateness. Recognition tasks assess identification when options are provided, whereas production requires retrieval with less support. Improvement on a controlled test should not be equated with fluent communicative use. Vocabulary research highlights noticing, reviewing, contextualising, and retrieval (Ghalebi

et al., 2020; Zarfsaz & Yeganehpour, 2021). Game-based learning organises practice around challenges, goals, feedback, and interaction. Language tasks may ask students to select meanings, match forms, complete sentences, or explain choices. Effective games align mechanics with learning goals rather than simply adding points or quizzes (Al Fatta et al., 2019; Escudeiro & Carvalho, 2013; Liu et al., 2020; Vandercruysse et al., 2012). Game-based outcomes vary with age, subject, design, duration, and assessment (Hartt et al., 2020; Karakoc et al., 2022). Plass et al. (2015) identify cognitive, motivational, affective, and sociocultural processes, including repeated exposure, retrieval, feedback, and peer explanation. Positive engagement is not itself evidence of vocabulary growth; achievement must be assessed separately. AI tools can generate contextual examples and feedback, but usefulness depends on accuracy, relevance, and clarity. Teachers should check meanings, collocations, grammar, and difficulty. Research on AI-assisted vocabulary learning reflects growing interest in generative systems while underscoring the need to examine actual learner interaction (Liu & Jin, 2025; Polyzi & Moussiades, 2023; Xu et al., 2025). Here, the teacher selected and reviewed materials; automatic difficulty adaptation was not verified. Three perspectives informed the study. Cognitive Load Theory highlights clear instructions and manageable steps (Sweller, 1988); collaborative learning emphasises explanation and shared responsibility (Admiraal et al., 2011; Stahl et al., 2006); and the Technology Acceptance Model identifies perceived usefulness and ease of use as relevant to acceptance (Davis, 1989). These perspectives guided planning but were not tested as causal mechanisms.

III. METHODOLOGY

3.1 Design and participants. A quantitative quasi-experimental pre-test–post-test design was used with two intact Grade 10 classes at Vinh Loc High School in Ho Chi Minh City. Convenience sampling was used because classes were already organised within the school timetable. The sample comprised 46 students, approximately 15–16 years old: 23 in the experimental group and 23 in the comparison group. Both classes were taught by the same teacher and followed the same curriculum. Students were not individually randomised, and each condition was represented by one intact class. The design enabled a practical classroom comparison but did not eliminate possible class-level differences. The intervention lasted six weeks, from Week 2 to Week 7, with three sessions per week and 18 regular English lessons per group. Both groups studied vocabulary from Units 1–3 of English 10: Global Success: Family Life, Humans and the Environment, and Music. Vocabulary objectives, lesson sequence, planned practice time, and homework demands were kept comparable. This was intended to reduce differences in curricular exposure, although it could not ensure identical attention or effort.

3.2 Instruments.

Vocabulary achievement was measured with tests covering 20 target words and phrases. Each test contained eight multiple-choice items, six matching items, and six contextual completion items with a word bank. Each correct response received 0.5 points, for a maximum of 10. The pre-test and post-test had similar structures and target vocabulary, but contexts, question order, and answer options varied. Students completed each test individually without learning aids in 30 minutes. The format mainly assessed recognition and controlled contextual use, not spontaneous speaking or extended writing. The experimental group completed a 30-item questionnaire after the post-test. Items used a five-point scale from 1 (strongly disagree) to 5 (strongly agree). Five positively worded items initially represented each of six domains: motivation, behavioural and cognitive engagement, enjoyment and emotional experience, perceived vocabulary learning effectiveness, collaboration, and technology acceptance. Only the experimental group completed the questionnaire; it described post-intervention perceptions, not changes from a baseline or differences between conditions. Positive wording may have encouraged acquiescence or socially desirable responses.

3.3 Procedure.

The researcher prepared Curipod questions, examples, and interactive slides before instruction. The teacher checked language, meanings, and instructions. Students received an orientation to joining lessons and submitting responses, using familiar vocabulary outside the assessed content. Experimental lessons began with a brief review or topic question, sometimes using a word cloud. New vocabulary was introduced through explanations, pronunciation practice, and examples. Students then completed recognition prompts, matching activities, and sentence exercises. The teacher reviewed answers and explained why particular words fitted their contexts. Selected sentence activities included AI-generated feedback. The teacher reviewed comments and clarified advice that was unclear or unsuitable. Where another attempt was available, students revised their responses. Individual tasks

were combined with small-team rounds in which students discussed meanings before submitting an answer. Some rounds awarded classroom points that did not affect school grades. Roles rotated to distribute participation. Lessons ended with short reviews, and printed prompts were available when internet interruptions prevented digital participation. The comparison group studied the same vocabulary through teacher explanation, textbook activities, and printed exercises. Students practised pronunciation, matched words with meanings, and completed sentences. The teacher checked answers and corrected errors. Pair discussion was used where appropriate, but lessons did not include Curipod, AI feedback, timed game rounds, or team points. A shared checklist supported consistency in content and planned practice. Lesson records noted vocabulary covered, activities completed, teaching time, attendance, and technical interruptions; they were not analysed as an independent measure of engagement.

3.4 Analysis and ethics.

Means and standard deviations were calculated using SPSS. Independent-samples t-tests compared baseline scores, post-test scores, and gain scores (post-test minus pre-test). Paired-samples t-tests examined within-group changes. Questionnaire analysis used descriptive statistics and Cronbach's alpha, with .70 as the stated reliability criterion. Item removal was exploratory. Student-level tests did not account for clustering within only two classes; p-values are therefore interpreted cautiously. Participants were informed about the study, participation was voluntary, and withdrawal carried no penalty. Participant codes protected confidentiality, and questionnaire responses did not affect grades.

IV. RESULTS AND DISCUSSION

4.1 Vocabulary achievement.

Before instruction, the comparison group mean was 7.413 (SD = 0.492), while the experimental group mean was 7.630 (SD = 0.505). The independent comparison was not significant, $t(44) = -1.479$, $p = .146$. The experimental group therefore began slightly higher, although the test did not detect a statistically significant baseline difference. This does not prove equivalence between intact classes, particularly given the small sample and absence of individual random assignment. After instruction, the comparison group mean was 7.304 (SD = 0.446), a mean change of −0.109 points. Its paired comparison was not significant, $t(22) = 1.045$, $p = .308$. The experimental group mean increased to 8.696 (SD = 0.670), a gain of 1.065 points; its paired comparison was significant, $t(22) = -8.662$, $p < .001$. The post-test difference favoured the experimental group by 1.3913 points, $t(44) = -8.294$, $p < .001$. Gain scores also differed, $t(44) = -7.287$, $p < .001$. The estimated difference in mean gains was 1.174 points, with a 95% confidence interval from 0.849 to 1.499 under the reported student-level model. Gain-score standard deviations were 0.499 in the comparison group and 0.590 in the experimental group, indicating variation within both classes. The pattern supports greater immediate improvement in the experimental class, subject to design limitations. It does not show that every student benefited or that AI alone caused the difference. The experimental condition combined teacher explanation, repeated encounters, interactive prompts, selected AI feedback, sentence practice, team discussion, and game-like challenges. The design cannot isolate the contribution of each component. Repeated opportunities to retrieve and reconsider target words provide one plausible explanation. Students encountered vocabulary through explanations and examples, then responded to recognition and matching prompts, completed contextual sentences, received feedback, and revisited difficult items. This sequence may have helped connect word forms and meanings with sentence contexts. Revising after feedback may also have made errors more visible. These interpretations are consistent with research on vocabulary strategies, contextual learning, and retention (Ghalebi et al., 2020; Zarfsaz & Yeganehpour, 2021), but retrieval processes and depth of word knowledge were not directly measured. The game-based structure may have supported participation by dividing practice into short tasks with clear goals. Research identifies potential value in feedback, challenge, and active involvement, while cautioning that outcomes depend on instructional design (Plass et al., 2015; Hartt et al., 2020). In this study, points were classroom incentives rather than grades, and the teacher remained responsible for explaining answers and checking content. The results relate to a teacher-mediated package, not competition or digital rewards in isolation. Cognitive load was not measured, so the study cannot claim that the activities reduced mental effort, even though short steps and clear instructions were intended to make tasks manageable (Sweller, 1988).

4.2 Student perceptions and questionnaire reliability.

All 23 experimental students completed the questionnaire. Item means ranged from 4.09 to 4.57. Domain means were 4.34 (SD = 0.461) for motivation, 4.35 (SD = 0.452) for enjoyment,

4.34 (SD = 0.432) for collaboration, and 4.40 (SD = 0.505) for technology acceptance. Responses were generally favourable. Interest in learning English had a mean of 4.57; comfort sharing ideas with teammates, 4.52; perceived usefulness, 4.52; ease of use, 4.48; and willingness to continue, 4.43. The pattern was not uniformly positive across activities. Willingness to review vocabulary at home had a lower mean of 4.13 than interest in classroom learning. Comfort sharing ideas with teammates (M = 4.52) was higher than perceived learning from peers (M = 4.09). Students may therefore enjoy interactive classroom practice without transferring the same effort to independent review or viewing peers as a major source of vocabulary learning. Enjoyment and acceptance can support participation, but are not substitutes for evidence of learning or independent study habits. Cronbach's alpha was .745 for motivation, .722 for enjoyment, and .782 for technology acceptance. Engagement increased from .669 to .697 after ENG2 was removed. Perceived effectiveness increased from .598 to .701 after EFF4 and EFF5 were removed. Collaboration remained at .697. Engagement and collaboration thus remained slightly below the stated .70 criterion, while some coefficients depended on exploratory item removal. The original five-item means should not be treated as scores for shortened scales. Retained effectiveness items concerning contextual understanding and confidence each averaged 4.26. Responses also had a restricted range: the minimum was three for most items and four for one item, and no disagreement response was recorded. This ceiling tendency limits the questionnaire's ability to capture moderate or critical experiences.

4.3 Implications and limitations.

Test results and questionnaire responses provide complementary but distinct evidence. Achievement data suggest a larger immediate gain in the experimental class, while the questionnaire describes students' experiences. Positive attitudes do not prove that students learned more, and test gains do not demonstrate increased motivation. The comparison group's small decline was not significant and should be interpreted as no detected improvement in this sample, not evidence that conventional teaching generally reduces vocabulary learning. The intervention fitted the normal curriculum and timetable, making it relevant to schools considering digital vocabulary practice. Teachers may use short interactive prompts after explicit presentation, move from recognition to contextual sentence work, and revisit difficult words in later lessons. Team tasks can be included when students have clear roles and time to explain choices. AI-generated examples and feedback should be reviewed before use; teachers should correct inaccurate meanings, clarify ambiguous comments, and ensure that suggestions suit learners' proficiency. Printed alternatives are useful when devices or internet access are unreliable. These recommendations reflect the need for instructional orchestration when adopting educational games and digital tools (Al Fatta et al., 2019; Escudeiro & Carvalho, 2013). Several limitations constrain conclusions. First, the sample came from one school and was selected through convenience sampling. With one intact class per condition, class characteristics were confounded with the intervention. Student-level tests did not account for clustering, so p-values are exploratory evidence within these classes rather than definitive estimates of a general treatment effect. Second, pilot testing suggested that the test forms may not have been equivalent. Uncertainty about form comparability affects interpretation of the magnitude of change. Third, no delayed post-test was administered, so retention and transfer to spontaneous speaking or extended writing cannot be established. Fourth, the questionnaire was completed only by the experimental group after instruction. All statements were positively worded, the sample was small, and the absence of disagreement responses suggests possible response bias or restricted variation. Reliability for engagement and collaboration was below the stated threshold, and two scales were modified during analysis. Fifth, cognitive load, peer-interaction quality, frequency of AI feedback, and changes to answers after feedback were not directly measured. The study also did not compare identical digital activities with and without AI feedback. The independent contribution of AI therefore cannot be separated from repeated practice, game structure, teacher guidance, and collaboration. Future research should include more classes and schools, use random assignment where feasible, and account for class-level clustering. It should establish test-form equivalence, add delayed assessments, and measure productive vocabulary in speaking and writing. Comparing identical digital activities with and without AI feedback would help isolate AI's contribution. Observations, interaction records, and interviews could explain how learners use feedback. Questionnaires should include balanced items and be validated with larger samples.

V. CONCLUSION

This study examined a six-week, teacher-guided Curipod intervention for Grade 10 vocabulary

learning at a high school in Ho Chi Minh City. The experimental class, which completed interactive prompts, contextual sentence tasks, selected AI-generated feedback, and team challenges, showed a larger immediate gain than the comparison class receiving conventional instruction. The difference in mean gains was 1.174 points in favour of the experimental group ($p < .001$ in the reported student-level analysis). Experimental students also reported generally favourable experiences, particularly in motivation, enjoyment, perceived usefulness, ease of use, and willingness to continue. The findings support cautious use of AI-supported game-based activities as one component of curriculum-aligned vocabulary instruction. They do not establish the independent effect of AI, prove long-term retention, or demonstrate transfer to unrestricted language use. The intervention depended on teacher selection of tasks, review of generated content, explanation of feedback, and management of classroom participation. Given the small sample, intact-class design, and measurement limitations, the results are promising classroom evidence that warrants replication. Future studies with more classes, validated parallel tests, delayed assessments, and comparisons that isolate AI feedback are needed before stronger claims can be made about effectiveness across Vietnamese secondary schools.

REFERENCES

Admiraal, W., Huizenga, J., Akkerman, S., & ten Dam, G. (2011). The concept of flow in collaborative game-based learning. Computers in Human Behavior, 27, 1185–1194.

Afzal, N. (2019). A study on vocabulary-learning problems encountered by BA English majors at the university level of education. Arab World English Journal, 10, 81–98. https://doi.org/10.24093/awej/vol10no3.6

Al Fatta, H., Maksom, Z., & Zakaria, M. (2019). Game-based learning and gamification: Searching for definitions. International Journal of Simulation: Systems, Science & Technology. https://doi.org/10.5013/IJSSST.a.19.06.41

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

Escudeiro, P., & Carvalho, C. (2013). Game-based language learning. International Journal of Information and Education Technology, 3, 643–647. https://doi.org/10.7763/IJIET.2013.V3.353

Ghalebi, R., Sadighi, F., & Bagheri, M. S. (2020). Vocabulary learning strategies: A comparative study of EFL learners. Cogent Psychology, 7.

Hartt, M., Hosseini, H., & Mostafapour, M. (2020). Game on: Exploring the effectiveness of game-based learning. Planning Practice & Research, 35(5), 589–604. https://doi.org/10.1080/02697459.2020.1778859

Karakoc, B., Eryılmaz, K., Özpolat, E., & Yıldırım, I. (2022). The effect of game-based learning on student achievement: A meta-analysis study. Technology, Knowledge and Learning, 27, 1–16. https://doi.org/10.1007/s10758-020-09471-5

Liu, Z., Shaikh, Z. A., & Gazizova, F. (2020). Using the concept of game-based learning in education. International Journal of Emerging Technologies in Learning, 15, 53–64.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. Educational Psychologist, 50, 258–283.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 409–426). Cambridge University Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Vandercruysse, S., Vandewaetere, M., & Clarebout, G. (2012). Game-based learning: A review on the effectiveness of educational games. https://doi.org/10.4018/978-1-4666-0149-9.ch032

Xu, S., Lo, C. K., & Ling, M. H. (2025). Leveraging generative AI in a 3D scenario-based game for vocabulary acquisition and conversation practice: Insights from the technology acceptance model. Cogent Education. https://doi.org/10.1080/2331186X.2025.2560059

Zarfsaz, E., & Yeganehpour, P. (2021). The impact of different context levels on vocabulary learning and retention. Shanlax International Journal of Education, 9, 24–34. https://doi.org/10.34293/education.v9i4.4049