Jurnal Onoma: Pendidikan, Bahasa dan Sastra, **Vol. 12, No. 3, 2026**

Investigating the Use of Kahoot for Long-Term Vocabulary Retention among Grade XI EFL Students at SMAN 12 Enrekang

Aisa Fitri[1]
Musdalifah[2]
Sri Rosmiana[3]
Ismail[4]
Umiyati Jabri[5]
[1]Bachelor Students at Muhammadiyah University of Enrekang, Indonesia
[2345]Muhammadiyah University of Enrekang, Indonesia

[1]aisafitri3112@gmail.com
[2]musdalifah25okt@gmail.com
[3]srirosmiana@gmail.com
[4]smileummaspul@gmail.com
[5]umiyatijabri@gmail.com

Abstract

Despite growing evidence supporting game-based learning in EFL vocabulary instruction, long-term retention effects of Kahoot remain largely underexplored, as most existing studies are limited to immediate post-test measurements, leaving its longitudinal impact on vocabulary retention empirically unverified. This study aimed to investigate the effectiveness of Kahoot-based instruction on long-term vocabulary retention and students' perceptions toward Kahoot among Grade XI EFL students at SMAN 12 Enrekang, Indonesia. Employing a quantitative quasi-experimental design with a non-equivalent control group, 48 students were divided into an experimental group (n = 24) receiving Kahoot-based instruction and a control group (n = 24) following rote learning methods across six instructional sessions. Data were collected through a 20-item vocabulary test administered at four measurement points, namely pre-test, post-test, and two delayed post-tests, and a 10-item Likert-scale questionnaire administered to the experimental group. Mixed ANOVA results revealed a significant Time * Group interaction effect (F(2.503, 115.145) = 3.277, p = .031, η_p^2 = .067), with the experimental group demonstrating significantly superior retention at the second delayed post-test (F(1, 46) = 5.126, p = .028, η_p^2 = .100), approximately five weeks after the completion of instruction, retaining a mean score of 62.29 compared to 45.62 in the control group. Students reported very positive perceptions toward Kahoot (overall M = 4.32), with enthusiasm (M = 4.71) and enjoyable learning atmosphere (M = 4.67) as the most valued features. These findings indicate that Kahoot-based instruction is significantly more effective than rote learning in sustaining long-term vocabulary retention, with practical implications for EFL teachers to integrate retrieval-based and game-based tools as a strategic approach to vocabulary instruction.

Keywords: *Kahoot; long-term vocabulary retention; EFL; Game-based learning; Quasi-experimental*

Introduction

English has established itself as the dominant global lingua franca, fundamentally shaping educational priorities worldwide. According to the British Council and Studyportals (2024), there are currently 183,262 English-taught programmes offered

globally, representing a 22% increase since 2021. This unprecedented global dominance creates urgent demands for English proficiency, particularly in English as a Foreign Language (EFL) contexts where learners must acquire language competency primarily through formal classroom instruction. Central to this proficiency is vocabulary, which serves as the fundamental building block of communication. As Nation (2013) emphasizes, without sufficient vocabulary, learners cannot convey or comprehend messages effectively. This is supported by Richards and Renandya (2002) and more recent studies (Hakami, 2025; Zeng et al., 2025), which highlight that vocabulary knowledge is the core component shaping overall language proficiency and academic achievement. Despite its central role, vocabulary learning in EFL contexts is often hindered by a recurring challenge: the problem of long-term retention. Many students memorize vocabulary for exams or quizzes but quickly forget the words when they are not revisited or used in meaningful contexts.

Despite its importance, EFL learners frequently face the challenge of long-term retention. Many students memorize vocabulary for immediate assessments but quickly forget it due to a lack of meaningful reinforcement (Rajayi & Maleki, 2023; Webb & Nation, 2017). Without reinforcement, vocabulary tends to fade rapidly from memory. To address this, Game-Based Learning (GBL) has emerged as a promising strategy. Prensky (2001) argues that games situate material in engaging contexts, while recent research confirms that GBL enhances motivation and critical thinking (Zhong, 2024). Among GBL tools, Kahoot is widely recognized for increasing participation and retention through competition and real-time feedback (Ahmed et al., 2022). This effectiveness is often attributed to Dual Coding Theory, which suggests that combining verbal and visual inputs enhances memory retention (Paivio, 1979).

However, a notable gap remained regarding Kahoot's long-term effects. Most existing studies were limited to immediate post-test measurements, leaving its longitudinal impact underexplored (Añora et al., 2025; Shofiana & Maemanah, 2025). Aisyah and Salmiah, (2024) specifically called for research into Kahoot's role in sustaining vocabulary retention across diverse educational settings. Furthermore, although evidence strongly supports Kahoot's effectiveness in boosting short-term vocabulary learning, a notable gap remained regarding its long-term effects, as Farhane (2025) included only one delayed post-test conducted three weeks after the treatment, while Ahmed et al. (2022) employed only one delayed post-test, whereas the present study included two delayed post-tests at different intervals to provide a more detailed measurement of long-term retention. This gap was particularly relevant in Indonesia, where exam-oriented instruction and limited language exposure already made long-term vocabulary retention a persistent challenge. At SMAN 12 Enrekang, preliminary observations during the "EDSA Goes to School" program in February 2025 revealed a heavy reliance on rote learning methods such as note-taking, memorization drills, and translation pair memorization. While students showed high enthusiasm for Kahoot during brief trials, its capacity to solve their long-term retention issues remained empirically unverified in this local context.

Therefore, this study investigated whether Kahoot-based instruction effectively promoted long-term retention among Grade XI students at SMAN 12 Enrekang. This study offered novelty in three key aspects. First, it represented the first research to examine Kahoot's long-term impact on vocabulary retention in the Indonesian EFL context, as previous research focused on other countries such as Morocco (Farhane, 2025) and Iran (Ahmed et al., 2022). Second, it employed two delayed post-tests administered at intervals of two and three weeks instead of one, providing more

comprehensive evidence of long-term retention. Third, it employed a quantitative approach by using both vocabulary test results and Likert-scale questionnaire data to measure students' perceptions toward the use of Kahoot. Specifically, this study addressed two research questions: (1) is there a significant difference in long-term vocabulary retention between students taught using Kahoot and those taught using rote learning; and (2) what are the students' perceptions toward using Kahoot for vocabulary learning?

Method

Research Design

This study employed a quantitative quasi-experimental design, specifically a non-equivalent control group design. According to Creswell and Creswell (2018), this approach is suitable for measuring the effect of an intervention on numerical data within intact classroom settings. The independent variable was the use of Kahoot as an interactive digital platform for vocabulary instruction, while the dependent variable was students' long-term vocabulary retention. The research compared an experimental group receiving Kahoot-based instruction with a control group following rote learning methods.

Participants

The study was conducted at SMAN 12 Enrekang in the academic year 2025/2026, with a sample consisting of 48 Grade XI students from two intact classes. One class was designated as the experimental group (n = 24) and the other as the control group (n = 24). To ensure parity, both groups were taught the same vocabulary materials derived from the national curriculum.

Research Instruments

The primary instrument was a 20-item vocabulary test consisting of multiple-choice and matching questions adapted from the English for Change textbook. To minimize memorization effects, question order and answer choices were randomized across all testing phases, with raw scores converted to a standardized 100-point scale. Additionally, a 10-item Likert-scale questionnaire (ranging from 1 to 5) was used to measure the experimental group’s perceptions toward Kahoot. To ensure quality, both instruments underwent content and construct validation by academic experts. Furthermore, a pilot study was conducted with a separate class to evaluate internal consistency and reliability. Reliability was assessed using the Cronbach’s Alpha coefficient, and item-total correlation analysis was performed to verify the validity of each individual item before the main data collection phase.

Data Collection Procedure

The data collection followed a longitudinal sequence to measure retention accurately. First, a pre-test was administered to both groups to establish baseline vocabulary knowledge. Second, both groups underwent six instructional sessions (treatments); the experimental group utilized Kahoot, while the control group followed rote learning methods. Third, an immediate post-test was conducted following the final session. Fourth, Delayed Post-test 1 (DP1) was administered two weeks after the post-test. Finally, Delayed Post-test 2 (DP2) was conducted three weeks after DP1 to measure

sustained retention, alongside the distribution of the perception questionnaire to the experimental group.

Data Analysis

Quantitative data were processed using SPSS through four stages: descriptive statistics, assumption tests, inferential hypothesis testing, and descriptive analysis of the questionnaire. Descriptive statistics (Mean and Standard Deviation) summarized score distributions across all testing phases. Before conducting the inferential analysis, normality (Shapiro-Wilk), homogeneity (Levene's Test), and sphericity (Mauchly's Test) were tested. If the assumption of sphericity was violated, appropriate corrections such as Greenhouse-Geisser or Huynh-Feldt were applied to adjust the degrees of freedom within the Mixed Analysis of Variance (Mixed ANOVA). This Mixed ANOVA was employed to examine the interaction effect between Time and Group, while post-hoc Bonferroni comparisons were used for pairwise analysis to identify specific differences between testing intervals. Finally, questionnaire data were interpreted using the five-point Likert scale classification proposed by Brown (2011).

Results

This section presents the findings of the study in response to the two research questions. Overall, the results indicated that Kahoot-based instruction produced significantly superior long-term vocabulary retention compared to rote learning, particularly at the five-week mark after instruction, and that students reported very positive perceptions toward the use of Kahoot in vocabulary learning (overall $M = 4.32$).

Descriptive Statistical Analysis

Descriptive statistical analysis was conducted to examine the distribution and general characteristics of students' vocabulary scores in both the experimental and control groups across four measurement points. The analysis included the mean, standard deviation, minimum, and maximum scores for each group at each testing phase.

Table 1. Descriptive Statistics of Vocabulary Scores Across All Testing Phases for Both Groups

	Treatment Group	Mean	Std. Deviation	Minimum	Maximum	*N*
Pre-test	Experimental	45.62	17.214	20	85	24
	Control	45.83	16.725	15	80	24
	Treatment Group	Mean	Std. Deviation	Minimum	Maximum	*N*
Post-test	Experimental	76.46	16.647	50	100	24
	Control	70.46	17.101	45	95	24
	Treatment Group	Mean	Std. Deviation	Minimum	Maximum	*N*
Delayed Post-test 1	Experimental	63.75	22.275	30	100	24
	Control	52.71	22.311	20	90	24
	Treatment Group	Mean	Std. Deviation	Minimum	Maximum	*N*
Delayed Post-test 2	Experimental	62.29	23.030	25	100	24
	Control	45.62	27.753	10	90	24

Note. M = Mean. SD = Standard Deviation.

As shown in Table 1, at the pre-test stage, both groups demonstrated comparable baseline vocabulary knowledge, with the experimental group obtaining a mean score of 45.62 (*SD* = 17.214, ranging from 20 to 85) and the control group achieving a mean of 45.83 (*SD* = 16.725, ranging from 15 to 80), confirming that both groups started from an equivalent level of initial knowledge before any treatment was given. Immediately following the six instructional sessions, both groups showed a substantial increase in scores at post-test, with the experimental group (*M* = 76.46, *SD* = 16.647, ranging from 50 to 100) outperforming the control group (*M* = 70.46, *SD* = 17.101, ranging from 45 to 95), indicating that the Kahoot method resulted in better learning outcomes than the rote learning method. The diverging trajectories became evident during the retention phases. At Delayed Post-test 1 (DP1), conducted two weeks after the post-test, the experimental group maintained a higher mean score (*M* = 63.75, *SD* = 22.275, ranging from 30 to 100) compared to the control group (*M* = 52.71, *SD* = 22.311, ranging from 20 to 90), suggesting that vocabulary retention was stronger in the group that received the Kahoot treatment. Most notably, at Delayed Post-test 2 (DP2), administered three weeks after DP1, the experimental group demonstrated remarkable stability (*M* = 62.29, *SD* = 23.030, ranging from 25 to 100), representing only a marginal decline of 1.46 points from their DP1 performance. In contrast, the control group's performance continued to decline sharply (*M* = 45.62, *SD* = 27.753, ranging from 10 to 90), a further drop of 7.09 points that nearly returned them to their initial pre-test baseline. These results confirmed that Kahoot-based instruction facilitated much higher and more sustained long-term retention compared to the rote learning method.

Assumption Tests

The normality of the data was examined using the Shapiro-Wilk test, which is considered more appropriate for small sample sizes (*n* = 24 per group). The results are presented in Table 2 below.

Table 2. Tests of Normality for Experimental and Control Groups

		Shapiro-Wilk		
	Treatment Group	Statistic	df	Sig.
Pre-test	Experimental	.947	24	.231
	Control	.976	24	.819
Post-test	Experimental	.929	24	.094
	Control	.917	24	.050
Delayed Post-test 1	Experimental	.931	24	.101
	Control	.939	24	.152
Delayed Post-test 2	Experimental	.936	24	.132
	Control	.893	24	.015

Note. S-W = Shapiro-Wilk. p > .05, meaning the data is normally distributed.

As presented in Table 2, the Shapiro-Wilk test results indicated that the majority of the data were normally distributed, as evidenced by significance values greater than .05. For the control group, the post-test scores yielded a significance value of exactly .050, which meets the standard minimum threshold for normality. At the delayed post-test 2, while the experimental group scores remained normally distributed (*SW* = .936, *p* = .132), the control group scores showed a slight deviation from normality (*SW* = .893, *p* = .015). Despite this single violation, Mixed ANOVA was still considered appropriate for the analysis, as the F-test in ANOVA is robust against minor deviations from normality

when group sample sizes are equal (n = 24 per group), as demonstrated through simulation studies by Schmider et al. (2010).

Prior to conducting the inferential analysis, the homogeneity of variance between the experimental and control groups was examined using Levene's Test of Equality of Error Variances. The results are presented in Table 3 below.

Table 3. Tests of Homogeneity of Variance for Experimental and Control Groups

		Levene Statistic	df1	df2	Sig.
Pre-test Score	Based on Mean	.145	1	46	.705
Post-test Score	Based on Mean	.032	1	46	.860
Delayed Post-test 1 Score	Based on Mean	.053	1	46	.818
Delayed Post-test 2 Score	Based on Mean	1.964	1	46	.168

As presented in Table 3, the results of Levene's Test indicated that the assumption of homogeneity of variance was satisfied across all four measurement points. Since all significance values exceeded the .05 threshold, ranging from the lowest at Delayed Post-test 2 (p = .168) to the highest at Post-test (p = .860), the error variance of the dependent variable was equal across both groups at every stage of the study.

Mauchly's Test of Sphericity was conducted to examine whether the variances of the differences between all combinations of measurement points were equal. The results are presented in Table 4 below.

Table 4. Mauchly's Test of Sphericity

					Epsilon[b]	
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	*df*	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Time	.645	19.641	5	.001	.773	.834	.333

Based on the results in Table 4, the assumption of sphericity for the variable Time was violated (Mauchly's W = .645, Approx. Chi-Square = 19.641, df = 5, p = .001). Consequently, the degrees of freedom required adjustment. Following Abdi (2010), the Huynh-Feldt correction (epsilon = .834) was preferred over the Greenhouse-Geisser correction (epsilon = .773) because the Greenhouse-Geisser estimate surpassed the .75 threshold, indicating that the Greenhouse-Geisser correction would be too conservative. Therefore, the Huynh-Feldt correction was applied throughout the subsequent Mixed ANOVA analysis.

Hypothesis Testing: Mixed ANOVA

To test the research hypothesis, a Mixed Analysis of Variance was conducted with Time (pre-test, post-test, delayed post-test 1, and delayed post-test 2) as the within-subjects factor and Treatment Group (experimental and control) as the between-subjects factor. Since the assumption of sphericity was violated, Huynh-Feldt corrected values were applied throughout the analysis. The results of the within-subjects effects are presented in Table 5 below.

Table 5. Tests of Within-Subjects Effects (Measure: Vocabulary Retention)

Source		Type III Sum of Squares	*df*	Mean Square	*F*	Sig.	η_p^2
Time	Huynh-Feldt	19479.521	2.503	7782.018	34.285	<.001	.427
Time * Group	Huynh-Feldt	1862.125	2.503	743.914	3.277	.031	.067
Error(Time)	Huynh-Feldt	26135.354	115.145	226.978			

Note. η_p^2 *= Partial Eta Squared.*

Based on the results in Table 5, the Mixed ANOVA revealed two crucial findings. First, there was a significant main effect of Time (F(2.503, 115.145) = 34.285, p < .001, η_p^2 = .427), indicating that the passage of time (the natural forgetting process) had a dominant 42.7 percent impact on the change in students' scores. Second, and most importantly, a significant interaction effect between Time and Group was observed, F(2.503, 115.145) = 3.277, p = .031, η_p^2 = .067. The significance level of p = .031 confirmed that the use of Kahoot successfully created a statistically different retention pattern compared to the rote learning method. Furthermore, the Partial Eta Squared value of η_p^2 = .067 demonstrated that the Kahoot treatment provided an additional 6.7 percent contribution to students' vocabulary retention. According to Cohen (1988), η_p^2 = .01 is classified as small, .06 as medium, and .14 as large; therefore, η_p^2 = .067 fell within the medium effect size category, indicating that the effect of Kahoot-based instruction was practically meaningful. While the natural process of forgetting (time) was a dominant factor with a large effect size (η_p^2 = .427), accounting for 42.7 percent of the total variance in students' retention scores, these results proved that Kahoot served as a significant retention booster, helping students maintain their vocabulary knowledge more effectively than the rote learning method. Regarding the between-subjects effect, the results are presented in Table 6 below.

Table 6. Tests of Between-Subjects Effects (Measure: Vocabulary Retention)

Source	Type III Sum of Squares	*df*	Mean Square	*F*	Sig.	η_p^2
Intercept	642412.688	1	642412.688	557.676	<.001	.924
Group	3366.750	1	3366.750	2.923	.094	.060
Error	52989.562	46	1151.947			

Note. η_p^2 *= Partial Eta Squared.*

Regarding the between-subjects effect presented in Table 6, the overall difference between the experimental and control groups across all time points did not reach statistical significance, F(1, 46) = 2.923, p = .094, η_p^2 = .060. This indicates that when the scores from all testing sessions are averaged together, the difference between the two groups is not statistically significant. This result is primarily attributable to the inclusion of the pre-test scores, where both groups demonstrated nearly identical baseline knowledge (mean = 45.62 for the experimental group and mean = 45.83 for the control group). However, this result should be interpreted in conjunction with the significant interaction effect (p = .031), which indicates that meaningful group differences emerged at specific time points rather than uniformly across all measurements.

Given the significant interaction effect, post-hoc pairwise comparisons with the Bonferroni adjustment were conducted to identify the specific sources of significant differences. The results are presented in Table 7 below.

Table 7. Pairwise Comparisons of Time for Both Groups (Bonferroni Adjustment)

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-27.729*	2.503	<.001	-34.630	-20.829
	3	-12.500*	3.057	.001	-20.928	-4.072
	4	-8.229	3.355	.108	-17.480	1.022
2	1	27.729*	2.503	<.001	20.829	34.630
	3	15.229*	2.854	<.001	7.360	23.098
	4	19.500*	2.884	<.001	11.548	27.452
3	1	12.500*	3.057	.001	4.072	20.928
	2	-15.229*	2.854	<.001	-23.098	-7.360
	4	4.271	2.004	.231	-1.255	9.797
4	1	8.229	3.355	.108	-1.022	17.480
	2	-19.500*	2.884	<.001	-27.452	-11.548
	3	-4.271	2.004	.231	-9.797	1.255

*Note. * The mean difference is significant at the .05 level. Adjustment for multiple comparisons: Bonferroni. Time 1 = Pre-test; Time 2 = Post-test; Time 3 = Delayed Post-test 1; Time 4 = Delayed Post-test 2.*

Based on the results of the Pairwise Comparisons with Bonferroni adjustment in Table 7, the analysis reveals a highly significant increase in vocabulary scores from the pre-test (Time 1) to the post-test (Time 2), with a mean difference of 27.729 and a significance level of $p < .001$. This finding demonstrates that the instructional intervention successfully provided an immediate and substantial impact on students' vocabulary mastery across both groups right after the material was taught. Furthermore, in the first delayed post-test (Time 3), students' scores still showed a significant difference compared to their initial pre-test scores, with a mean difference of 12.500 ($p = .001$), indicating that in the short term, students generally remained capable of retaining a significant portion of the learned vocabulary. However, a notable decline in memory retention was observed by the second delayed post-test (Time 4), where the difference in scores compared to the initial pre-test became statistically non-significant (mean difference = 8.229, $p = .108$), suggesting that over a longer period without further reinforcement, the learning effect began to fade. Overall, the comparisons between the post-test and both delayed post-tests show a consistent and statistically significant decline in scores ($p < .001$), confirming a gradual forgetting process occurring among all students.

The Univariate Tests for between-group comparisons at each specific testing point are presented in Table 8 below.

Table 8. Univariate Tests for Between-Group Comparisons at Each Time Point

Time		Sum of Squares	df	Mean Square	F	Sig.	η_p^2
1	Contrast	.521	1	.521	.002	.966	.000
	Error	13248.958	46	288.021			
2	Contrast	432.000	1	432.000	1.517	.224	.032
	Error	13099.917	46	284.781			
3	Contrast	1463.021	1	1463.021	2.944	.093	.060
	Error	22861.458	46	496.988			

4	Contrast	3333.333	1	3333.333	5.126	.028	.100
	Error	29914.583	46	650.317			

Note. η_p^2 *= Partial Eta Squared. DP 1 = Delayed Post-test 1; DP 2 = Delayed Post-test 2. Each F tests the simple effects of Treatment Group within each level of the other effects shown.*

Table 8 presents the results of the Univariate Tests, which analyze the differences between the experimental group using Kahoot and the control group using rote learning at each specific testing point. At Time 1 (Pre-test), the results show no significant difference between the two groups ($F(1, 46) = .002$, $p = .966$, $\eta_p^2 = .000$), confirming that both groups started with equivalent initial knowledge. At Time 2 (Post-test), the difference between the groups remained non-significant ($F(1, 46) = 1.517$, $p = .224$, $\eta_p^2 = .032$), indicating that the Kahoot method accounted for only 3.2 percent of the variance in students' scores right after instruction ended. At Time 3 (Delayed Post-Test 1), the difference between the groups began to emerge but had not yet reached statistical significance ($F(1, 46) = 2.944$, $p = .093$, $\eta_p^2 = .060$), yet the effect size had grown to 6 percent, indicating that Kahoot's influence on retention was increasing. A significant difference finally appeared at Time 4 (Delayed Post-Test 2), with $F(1, 46) = 5.126$ and $p = .028$, and the Partial Eta Squared value increased to .100, signifying that using Kahoot accounted for 10 percent of the variance in students' scores at the five-week mark. This progressive intensification from 3.2 percent at post-test to 6.0 percent at Delayed Post-Test 1 and to 10.0 percent at Delayed Post-Test 2 demonstrates that while the rote learning group's memory faded more rapidly, the Kahoot group was significantly more successful in maintaining vocabulary knowledge over time.

Based on the complete statistical results, the Mixed ANOVA revealed a significant interaction effect between Time and Treatment Group ($F(2.503, 115.145) = 3.277$, $p = .031$, $\eta_p^2 = .067$), and the Univariate Tests confirmed a statistically significant between-group difference at Delayed Post-Test 2 ($F(1, 46) = 5.126$, $p = .028$, $\eta_p^2 = .100$). Since both significance values fell below the alpha level of .05, the null hypothesis (H0) is rejected and the alternative hypothesis (H1) is accepted. It can therefore be concluded that the use of Kahoot significantly affected students' long-term vocabulary retention compared to rote learning methods. The visual representation of these score trends and the diverging retention trajectories between the two groups can be observed in the profile plot presented in Figure 1 below.

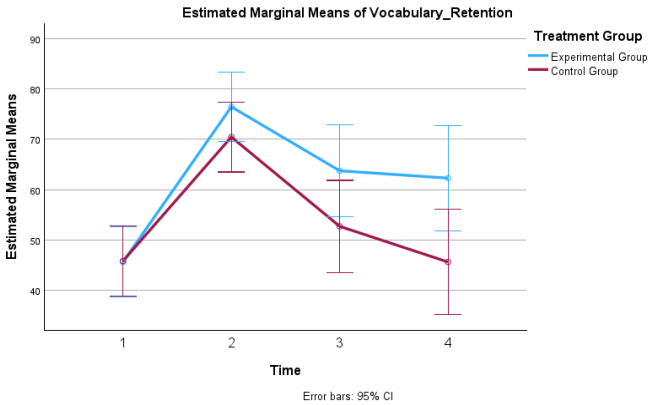

Figure 1. Profile Plot Of The Estimated Marginal Means

Note: Time 1 = Pre-test; Time 2 = Post-test; Time 3 = Delayed Post-test 1; Time 4 = Delayed Post-test 2

Descriptive Analysis of Questionnaire Data

In addition to the vocabulary test data, this study employed a Likert-scale questionnaire to measure students' perceptions toward the use of Kahoot in vocabulary learning. The questionnaire consisted of 10 statements rated on a five-point scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), administered to all 24 students of the experimental group after the completion of Delayed Post-Test 2. Descriptive statistical analysis was conducted to examine students' responses for each questionnaire item, and mean scores were interpreted using the classification proposed by Brown (2011). The results are presented in Table 9 below.

Table 9. Descriptive Statistical Analysis of Questionnaire Data

Item	*N*	*Min*	*Max*	*Mean*	*Std. Dev.*	*Category*
Q1	24	4	5	4.71	0.464	Very Positive
Q2	24	3	5	4.29	0.751	Very Positive
Q3	24	3	5	4.38	0.711	Very Positive
Q4	24	3	5	4.25	0.737	Very Positive
Q5	24	3	5	4.17	0.816	Positive
Q6	24	3	5	4.38	0.711	Very Positive
Q7	24	3	5	4.00	0.834	Positive
Q8	24	3	5	4.08	0.654	Positive
Q9	24	3	5	4.67	0.565	Very Positive
Q10	24	3	5	4.25	0.794	Very Positive
Overall	24	3	5	4.32	0.715	Very Positive

Note.

a) *Classification based on Brown (2011): 4.21 to 5.00 = Very Positive; 3.41 to 4.20 = Positive; 2.61 to 3.40 = Neutral; 1.81 to 2.60 = Negative; 1.00 to 1.80 = Very Negative.*

b) Percentages (%) are calculated by: (Frequency / *n*) x 100, where *n* = 24

Table 9 presents the mean scores and categories for each questionnaire item, with results ranging from 4.00 to 4.71. Specifically, the highest mean score was recorded for Q1 (M = 4.71), where students expressed that learning vocabulary using Kahoot made them more enthusiastic than rote learning methods. In contrast, the lowest mean score was found in Q7 (M = 4.00), which addressed whether students still remembered vocabulary from Kahoot after a certain period. Despite this variation, the overall mean score of 4.32 indicates that students' general perception toward using Kahoot was Very Positive based on Brown's (2011) criteria.

Notably, none of the students selected "Strongly Disagree" or "Disagree" for any item, indicating a complete absence of negative perceptions toward Kahoot. The highest level of enthusiasm was observed in Q1 and Q9, where 17 students (70.8%) selected "Strongly Agree" for both items. For Q1, which directly compared Kahoot to rote learning, the remaining 7 students (29.2%) selected "Agree," resulting in a 100 percent positive response rate. This confirms that the gamified nature of Kahoot was highly effective in fostering enthusiasm and an enjoyable learning environment.

While the majority of responses were positive, some students maintained a Neutral stance on specific items. This was most visible in Q7 regarding long-term memory, where 8 students (33.3%) chose the neutral option, and in Q10 regarding future use, where 5 students (20.8%) remained neutral. Despite these neutral stances, the combined positive response (Agree and Strongly Agree) remained dominant, reaching 66.7 percent for Q7 and 79.1 percent for Q10. These findings suggest that while a few

students were cautious about long-term effects, the overwhelming majority perceived Kahoot as a superior and motivating tool for vocabulary learning.

Discussion

The Effect of Kahoot on Long-Term Vocabulary Retention

The first research question asked whether Kahoot-based instruction significantly affected students' long-term vocabulary retention compared to rote learning. Based on the Mixed ANOVA results, the question was answered affirmatively. A significant Time x Group interaction effect was confirmed ($F(2.503, 115.145) = 3.277$, $p = .031$, $\eta_p^2 = .067$), demonstrating that the experimental group and the control group followed meaningfully different retention trajectories across the five-week measurement period. This interaction effect constituted the central statistical evidence supporting the alternative hypothesis (H1) and leading to the rejection of the null hypothesis (H0). More specifically, a statistically significant between-group difference emerged at the second delayed post-test ($F(1, 46) = 5.126$, $p = .028$, $\eta_p^2 = .100$), at which point the experimental group retained a mean score of 62.29 while the control group had declined to 45.62, producing a gap of 16.67 points approximately five weeks after instruction concluded.

The trajectory of scores across all four measurement points confirmed the hypothesis in a nuanced and informative way. Both groups began at statistically equivalent baseline levels at pre-test, with the experimental group at $M = 45.62$ ($SD = 17.214$) and the control group at $M = 45.83$ ($SD = 16.725$), and both improved substantially at post-test, with the experimental group reaching $M = 76.46$ ($SD = 16.647$) and the control group reaching $M = 70.46$ ($SD = 17.101$). However, from post-test onward, their trajectories diverged. The experimental group's scores declined from 76.46 at post-test to 63.75 at the first delayed post-test (DP1), and then to 62.29 at the second delayed post-test (DP2), a further reduction of only 1.46 points across three additional weeks. The control group, by contrast, declined from 70.46 at post-test to 52.71 at DP1 and continued falling to 45.62 at DP2, a further reduction of 7.09 points, nearly returning to its pre-test baseline of 45.83. This diverging pattern is precisely what a significant Time x Group interaction represents: the two groups did not simply differ in average score but in the rate and direction of change in their retention over time.

These findings were consistent with and extended the results of several previous studies. Ahmed et al. (2022) conducted a quasi-experimental study with 50 Iranian intermediate EFL learners divided into a Kahoot-based experimental group and a traditional instruction control group. Their results showed that the experimental group significantly outperformed the control group on both the immediate post-test and a delayed post-test administered three weeks after instruction. Farhane (2025) reported parallel findings with Moroccan EFL secondary school students, where the Kahoot-instructed group achieved significantly higher mean scores on both immediate and delayed assessments, with the delayed post-test conducted three weeks after the intervention. The present study extended these findings in two important respects. First, unlike Ahmed et al. (2022) and Farhane (2025), who each employed only one delayed post-test, this study administered two delayed measurements at two-week and five-week intervals, providing a more detailed longitudinal picture of the retention trajectory. Second, this design revealed that the between-group difference was not yet statistically significant at DP1 ($p = .093$) but became significant at DP2 ($p = .028$), indicating that Kahoot's retention advantage intensified progressively rather than

appearing immediately, a pattern that a single delayed post-test design would not have been capable of detecting.

The Mixed ANOVA also revealed a large main effect of Time ($F(2.503, 115.145) = 34.285$, $p < .001$, $\eta_p^2 = .427$), indicating that natural forgetting accounted for 42.7 percent of the total variance in retention scores across all participants regardless of group. This outcome was theoretically anticipated and aligned with Ebbinghaus's (1885) Forgetting Curve, which established that memory traces decay most rapidly in the period immediately following initial learning before gradually stabilizing. In the present study, this pattern was visible in both groups' score trajectories from post-test to DP1. The critical distinction was that the experimental group demonstrated a substantially diminished decline between DP1 and DP2, dropping by only 1.46 points, while the control group continued to decline by 7.09 points across the same interval. This pattern in the experimental group suggested that Kahoot-based instruction had facilitated the consolidation of vocabulary into more stable long-term memory representations, consistent with the Multi-Store Memory Model proposed by Atkinson and Shiffrin (1968), which posited that rehearsal and repeated processing of information facilitated its transfer from short-term to long-term memory.

Three cognitive mechanisms embedded within Kahoot's instructional design appear to account for why it produced superior long-term retention compared to rote learning. First, Retrieval Practice Theory (Roediger & Butler, 2011) established that actively retrieving information from memory strengthened memory traces more effectively than passive re-exposure or repetition. In the present study, each of the six Kahoot sessions required students to actively recall target vocabulary under time pressure during competitive quiz interactions, providing distributed retrieval practice across multiple sessions. This form of repeated active retrieval is precisely what the theory consistently associates with more durable long-term retention. Second, Dual Coding Theory (Paivio, 1979) proposed that presenting information through both verbal and visual channels simultaneously activated two independent cognitive systems, creating stronger and more diverse memory traces than verbal-only instruction. Kahoot's image-supported questions operationalized this principle by coupling word-meaning associations with visual stimuli during each quiz session, meaning that target vocabulary was encoded through two independent memory channels rather than one. Third, Kahoot delivered immediate corrective feedback after every quiz item, enabling students to identify errors and consolidate accurate lexical knowledge in real time, thereby preventing incorrect encodings from becoming stabilized in long-term memory (Roediger & Butler, 2011).

By contrast, the rote learning method used by the control group relied primarily on repetitive memorization through oral question-and-answer exchanges and written exercises on the board. While this approach involved some degree of verbal retrieval, it lacked the systematic competitive retrieval pressure, immediate corrective feedback, and visual encoding that Kahoot provided to the experimental group. As Sweller (1988) noted in Cognitive Load Theory, learning activities that do not promote meaningful processing of information are less effective at building durable long-term memory representations, regardless of the effort invested in repetition. The theoretical contrast between Kahoot's multi-mechanism approach and the control group's conventional method was directly reflected in the empirical contrast observed in the two groups' retention trajectories, particularly at DP2.

The dominant time effect should not be interpreted as evidence that Kahoot's contribution was negligible. The Partial Eta Squared value of $\eta_p^2 = .067$ represented the

proportion of retention variance specifically and uniquely attributable to the instructional method after statistically controlling for the universal effects of time that operated across both groups. According to Cohen (1988), a value of η_p^2 = .067 falls within the medium effect size category. A medium effect size produced by a bounded six-session classroom intervention operating against the powerful and unavoidable force of natural forgetting is educationally meaningful and practically significant. Furthermore, the effect size observed specifically at DP2 (η_p^2 = .100) approached the large threshold, reinforcing the practical importance of Kahoot's advantage at the final measurement point. This finding is corroborated by Özdemir's (2024) comprehensive meta-analysis of 43 independent studies, which confirmed that Kahoot exerted a very large positive effect on knowledge retention across diverse educational contexts.

Students' Perceptions of Kahoot as a Vocabulary Learning Tool

The second research question asked what students' perceptions were toward Kahoot as a vocabulary learning tool. The questionnaire results revealed that the experimental group held consistently very positive perceptions across all ten measured dimensions, with an overall mean score of 4.32 on a five-point Likert scale. This value fell within the Very Positive category according to (Brown, 2011) classification. Notably, no participant selected Disagree or Strongly Disagree on any of the ten items, indicating a uniformly positive reception. This pattern carried particular weight given that students had used Kahoot across six complete instructional sessions rather than encountering it for the first time, suggesting that their positive perceptions reflected genuine and sustained experience with the platform rather than novelty effects.

These findings were consistent with perception results reported across diverse EFL contexts internationally. Ji (2025) found that 91.9 percent of Korean university EFL students considered Kahoot effective for vocabulary learning, with students reporting that the platform increased their motivation and engagement. Farhane (2025) reported that all interviewed Moroccan EFL secondary school participants perceived Kahoot as making vocabulary lessons more enjoyable and engaging than conventional methods. Mubarok et al. (2025) similarly documented high levels of student motivation and enthusiasm in Indonesian senior high school EFL classrooms where Kahoot was used for vocabulary instruction. The present study extended these findings to the Indonesian Grade XI secondary school context, indicating that positive perceptions of Kahoot were not specific to any single national or institutional setting but appeared consistent across varied EFL environments.

The two highest-rated items were Q1, which measured student enthusiasm toward Kahoot sessions relative to rote learning (M = 4.71, SD = 0.464), and Q9, which assessed the extent to which Kahoot generated an enjoyable learning atmosphere (M = 4.67, SD = 0.565). On both items, 70.8 percent of students selected Strongly Agree and the remaining 29.2 percent selected Agree, producing a fully positive distribution with no neutral or negative responses whatsoever. Self-Determination Theory (Ryan & Deci, 2000) proposed that educational activities fulfilling learners' psychological needs for autonomy, competence, and relatedness generated authentic motivation that supported deeper cognitive engagement with learning content. In the Kahoot sessions, students exercised autonomy through independent answer selection, experienced competence through real-time score feedback after each item, and fulfilled their need for relatedness through competitive peer interaction. The very positive responses on Q1 and Q9

indicated that these three psychological needs were consistently met across the six sessions.

Students also responded positively to Kahoot's immediate feedback mechanism. Q6, which measured whether the correct or incorrect answer display helped students immediately identify their mistakes, received a mean score of 4.38 (*SD* = 0.711), with 87.5 percent of students selecting Agree or Strongly Agree. This result provided student-level confirmation of a mechanism identified as theoretically essential for long-term retention by Roediger and Butler (2011): feedback delivered immediately after a retrieval attempt enabled learners to correct errors and consolidate accurate vocabulary knowledge in real time.

Perceptions of Kahoot's image-based question format were also positive overall. Q5, which measured whether the image feature in Kahoot made vocabulary easier to remember, received a mean score of 4.17 (*SD* = 0.816), falling within the Positive category according to Brown (2011). It produced the highest neutral response rate among the cognitively focused items, with 25.0 percent selecting Neutral, 41.7 percent selecting Agree, and 33.3 percent selecting Strongly Agree. This variation is interpretable within the framework of Dual Coding Theory (Paivio, 1979), which posits that dual encoding occurs automatically when both verbal and visual cognitive channels are simultaneously activated, regardless of whether individual learners explicitly perceive the visual dimension of the experience as the primary driver of easier remembering.

The most analytically nuanced finding in the questionnaire data appeared in Q7, which asked directly whether students still remembered vocabulary learned through Kahoot after some time had passed (*M* = 4.00, SD = 0.834). Unlike all other items, Q7 produced an almost equal distribution across three response categories: 33.3 percent Strongly Agree, 33.3 percent Agree, and 33.3 percent Neutral. This distribution is interpretable as metacognitive accuracy rather than dissatisfaction: students who selected Neutral likely recognized that some degree of forgetting had occurred naturally, which is precisely what the objective retention data confirmed, with the experimental group's mean declining from 76.46 at post-test to 63.75 at DP1 and 62.29 at DP2. This realistic self-assessment strengthened rather than undermined the credibility of the overall questionnaire findings. The fact that 66.6 percent of students still responded positively on Q7 indicated that they recognized a meaningful net retention benefit despite acknowledging that complete preservation of vocabulary knowledge was not achieved.

Synthesis: Explaining Retention Outcomes through Student Perceptions

The findings of the two research questions are most meaningfully understood in relation to each other rather than in isolation. The significant Time x Group interaction in the retention data did not arise from any single mechanism but from four cognitive and affective mechanisms that operated simultaneously within each Kahoot session, each of which was directly reflected and confirmed by the questionnaire data. First, Kahoot generated sustained affective engagement across all six sessions, as reflected in Q1 (*M* = 4.71) and Q9 (*M* = 4.67). Zhong (2024) established that emotional engagement in game-based learning deepened vocabulary encoding by creating affective associations between words and the enjoyable context in which they were encountered, and that these associations slowed subsequent forgetting. Second, six repeated quiz sessions provided distributed retrieval practice, as reflected in Q2 (*M* = 4.29) and Q3 (*M* = 4.38), ensuring that vocabulary was actively recalled multiple times rather than encountered

passively. Roediger and Butler (2011) demonstrated that this form of repeated active retrieval produced significantly more durable memory traces than passive repetition. Third, image-supported quiz questions facilitated dual encoding, as reflected in Q5 (M = 4.17), activating both verbal and visual memory systems simultaneously and creating stronger memory representations (Paivio, 1979). Fourth, immediate corrective feedback after every item, as reflected in Q6 (M = 4.38), enabled students to consolidate accurate vocabulary knowledge in real time, preventing incorrect encodings from becoming entrenched (Roediger & Butler, 2011). These four mechanisms operating simultaneously across six sessions explained why Kahoot's effect size grew progressively from 3.2 percent at post-test to 6.0 percent at DP1 and to 10.0 percent at DP2.

Conclusion

This study investigated the effect of Kahoot-based instruction on long-term vocabulary retention and students' perceptions toward Kahoot among Grade XI EFL students at SMAN 12 Enrekang. Based on the statistical analyses conducted using SPSS version 31 and the Mixed ANOVA framework, two conclusions are drawn in accordance with the research questions.

First, the use of Kahoot significantly affected students' long-term vocabulary retention compared to rote learning methods. While both groups demonstrated comparable short-term gains at the post-test, with the experimental group reaching M = 76.46 and the control group M = 70.46, the experimental group consistently maintained higher vocabulary scores than the control group at both delayed measurement points. The between-group difference became statistically significant at Delayed Post-Test 2 ($F(1, 46) = 5.126$, $p = .028$, $\eta_p^2 = .100$), with the experimental group retaining a mean score of 62.29 compared to 45.62 in the control group, five weeks after the completion of instruction. The significant Time x Group interaction ($F(2.503, 115.145) = 3.277$, $p = .031$, $\eta_p^2 = .067$) further confirmed that the two groups followed meaningfully different retention trajectories over time. These findings indicate that Kahoot-based instruction not only supported immediate vocabulary learning but also produced more durable memory traces compared to rote learning methods, leading to the rejection of the null hypothesis (H0) and the acceptance of the alternative hypothesis (H1).

Second, the experimental group held a very positive overall perception toward the use of Kahoot for vocabulary learning, with an overall mean score of 4.32 on a five-point Likert scale, classified as Very Positive according to Brown's (2011) criteria. Students perceived Kahoot as an engaging, motivating, and effective tool for vocabulary learning across all dimensions measured, including enthusiasm (Q1, M = 4.71), enjoyable learning atmosphere (Q9, M = 4.67), and immediate feedback utility (Q6, M = 4.38). These perceptions are consistent with the objective retention data, suggesting that the affective and cognitive engagement facilitated by Kahoot collectively contributed to the superior long-term vocabulary retention observed in the experimental group.

The findings of this study have several implications. Theoretically, the results support and extend Retrieval Practice Theory (Roediger & Butler, 2011), Dual Coding Theory (Paivio, 1979), and Self-Determination Theory (Ryan & Deci, 2000) in the EFL vocabulary learning context. Practically, the findings demonstrate that Kahoot is not merely an entertainment tool but a pedagogically effective platform whose features directly support the cognitive processes associated with long-term vocabulary retention. For schools, it is suggested to provide adequate technological infrastructure to support

the integration of game-based learning platforms. For teachers, Kahoot should be utilized as a strategic tool to maintain student motivation and support retrieval-based vocabulary practice, ideally in combination with other methods for more intensive practice. For future researchers, it is suggested to replicate this study with larger and more diverse samples across different school contexts and proficiency levels in Indonesia, to extend the retention interval beyond five weeks, and to investigate the combination of Kahoot with other innovative tools as a promising direction for future research.

Acknowledgment

The author expresses sincere gratitude to supervisors Musdalifah, S.Pd., M.Pd., and Hj. Sri Rosmiana, S.Pd., M.Pd., for their patient guidance and invaluable support throughout the completion of this research. Special thanks are also extended to examiners Dr. Ismail, S.Pd., M.Pd., and Dr. Hj. Umiyati Jabri, S.Pd., M.Pd., for their constructive feedback and suggestions that significantly improved this research. Gratitude is also extended to the Principal of SMAN 12 Enrekang, Suardam Djamadi, S.Pd., M.M., for generously granting research permission, and to the English teacher Mrs. Wahda for her warm cooperation. The author is also grateful to all Grade XI A and XI B students at SMAN 12 Enrekang for their enthusiastic cooperation and participation throughout the data collection process.

References

Abdi, H. (2010). Greenhouse–Geisser correction. In N. J. Salkind (Ed.), Encyclopedia of research design (pp. 545–548). Sage Publications. https://doi.org/10.4135/9781412961288.n168

Ahmed, A. A. A., Sayed, B. T., Wekke, I. S., Widodo, M., Rostikawati, D., Ali, M. H., Abdul Hussein, H. A., & Azizian, M. (2022). An empirical study on the effects of using Kahoot as a game-based learning tool on EFL learners' vocabulary recall and retention. Education Research International, 2022, 1–10. https://doi.org/10.1155/2022/9739147

Aisyah, N., & Salmiah, M. (2024). The effect of Kahoot as a learning media on students' vocabulary mastery. Indonesian EFL Journal, 10(2), 197–204. https://doi.org/10.25134/ieflj.v10i2.10081

Añora, A. R. I., Jose, A. E. S., & Fernandez, E. J. (2025). The effect of using Kahoot as a game-based learning tool on vocabulary acquisition of grade 10 learners. European Journal of Literature, Language and Linguistics Studies, 9(1), 100-113. https://doi.org/10.46827/ejlll.v9i1.608

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. Psychology of Learning and Motivation, 2, 89–195. https://doi.org/10.1016/S0079-7421(08)60422-3

British Council, & Studyportals. (2024). Mapping English-taught programmes worldwide. British Council. https://takeielts.britishcouncil.org/sites/default/files/mapping_english-taught_programmes_worldwide_0.pdf

Brown, J. D. (2011). Likert items and scales of measurement. Shiken: JALT Testing & Evaluation SIG Newsletter, 15(1), 10–14. https://teval.jalt.org/test/PDF/Brown34.pdf

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE publications.

Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology. (H. A. Ruger & C. E. Bussenius, Trans.). Teachers College, Columbia University. (Original work published 1885)

Farhane, H. (2025). An investigation of the effect of Kahoot on vocabulary retention among Moroccan EFL secondary school students: A quasi-experimental study. Journal of English Language Teaching and Applied Linguistics, 7(3), 242-257. https://doi.org/10.32996/jeltal.2025.7.3.25

Hakami, A. (2025). The influences of the English language vocabulary on EFL learners' English language competency proficiency. International Journal of Language and Literary Studies, 7(2), 431–447. https://doi.org/10.36892/ijlls.v7i2.2090

Ji, E. K. (2025). Students' perceptions of Kahoot! for vocabulary learning from various lenses. Studies in English Education, 30(4), 19–57. https://doi.org/10.22275/see.30.4.02

Mubarok, D. I. R., Bilqis, M., & Imamyartha, D. (2025). Gamified English instruction mediated by kahoot! at senior high school: evidence for vocabulary gains. IJIET (International Journal of Indonesian Education and Teaching), 9(1), 113–126. https://doi.org/10.24071/ijiet.v9i1.10435

Nation, I. S. P. (2013). Learning Vocabulary in Another Language (2nd ed.). Cambridge: Cambridge University Press.

Özdemir, O. (2024). Kahoot! Game-based digital learning platform: A comprehensive meta-analysis. Journal of Computer Assisted Learning, 41(1), e13084. https://doi.org/10.1111/jcal.13084

Paivio, A. (1979). Imagery and Verbal Processes. Psychology Press. https://doi.org/10.4324/9781315798868

Prensky, M. (2001). Digital Game-Based Learning. McGraw-Hill.

Rajayi, S., & Maleki, R. (2023). On the impact of teaching vocabulary through asynchronous social media on EFL learners' vocabulary learning and retention. Anatolian Journal of Education, 8(2), 181–194. https://doi.org/10.29333/aje.2023.8212a

Richards, J. C., & Renandya, W. A. (Eds.). (2002). Methodology in language teaching: An anthology of current practice. Cambridge University Press.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. Trends in Cognitive Sciences, 15(1), 20-27. https://doi.org/10.1016/j.tics.2010.09.003

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American psychologist, 55(1), 68-78.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 6(4), 147–151. https://doi.org/10.1027/1614-2241/a000016

Shofiana, K. L., & Maemanah, A. (2025). Exploring the use of Kahoot as a game-based learning tools in vocabulary teaching to senior high school students: A

systematic review. Jurnal Penelitian Ilmu Pendidikan Indonesia, 4(2), 893–901. https://jpion.org/index.php/jpi

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

Webb, S., & Nation, P. (2017). How vocabulary is learned. Oxford University Press.

Zeng, Y., Kuo, L. J., Chen, L., Lin, J. A., & Shen, H. (2025). Vocabulary instruction for English learners: A systematic review connecting theories, research, and practices. Education Sciences, 15(3), Article 262. https://doi.org/10.3390/educsci15030262

Zhong, K. (2024). The effect of game-based vocabulary learning on middle school students' motivation and vocabulary retention. Proceedings of ICILLP 2024 Workshop, 166-172. https://doi.org/10.54254/2753-7048/74/2024.BO17947