

Integrating Duolingo in EFL Classrooms: A Gamified Approach to Strengthening Vocabulary and Reading Competence

¹Bima Samudra Aji Pratama*, ²Yuri Lolita

^{1,2}Universitas Negeri Surabaya, Indonesia

*Corresponding Author

Email: bima.21073@mhs.unesa.ac.id

Abstract

This study investigates the impact of gamified learning through Duolingo on the vocabulary mastery of EFL students and its subsequent effect on reading Skill. Grounded in theories of gamification, self-regulated learning (SRL), and the Octalysis Framework, the research employed a quasi-experimental design involving 62 eleventh-grade students at one Senior High School at Mojokerto, East Java, divided into an experimental group using Duolingo and a control group taught with conventional methods. A vocabulary test consisting of 25 items across five reading-based indicators was administered as both pre-test and post-test. Several statistics step were applied to analyze the data. Further analysis across five indicators revealed that the most notable gains were in students' ability to recognize word meanings in context and apply vocabulary in varied reading tasks. These findings highlight the pedagogical value of Duolingo in improving not only vocabulary acquisition but also its application in reading skill. Gamified language learning platforms are thus recommended as effective tools to support vocabulary development in EFL contexts.

Keywords: Duolingo, EFL Learners, gamification, reading, vocabulary

INTRODUCTION

Vocabulary is one of the most important parts of learning English because it is the foundation for developing other language skills. Without enough vocabulary, students will find it hard to understand reading texts, follow spoken conversations, or express their ideas clearly. (Nation, 2022) explains that vocabulary supports all four skills: speaking, listening, reading, and writing and that students cannot communicate well in a second language without a strong vocabulary base. In Indonesia, English is mostly learned in the classroom, so vocabulary learning is even more important. (Abdullah Alzahrani et al., 2020) also found that vocabulary size has a strong effect on reading comprehension, especially for EFL learners who have limited exposure to English outside of school. However, many senior high school students in Indonesia still struggle with vocabulary, making it harder for them to read and understand academic texts.

One strategy that has become popular in recent years is gamification, which means adding game-like elements—such as points, badges, levels, and progress tracking, to learning activities to make them more interesting and motivating (Deterding et al., 2011). In language learning, gamification can turn simple vocabulary practice into an interactive experience. Research has shown that gamified activities can help students remember vocabulary better, reduce anxiety, and encourage independent practice (Al-Dosakee & Ozdamli, 2021). Jane et al. (2024) also found that gamified learning can give EFL students a low-pressure space where they can practice and make mistakes without feeling stressed.

Duolingo is one of the most popular gamified language learning apps today. It teaches vocabulary and grammar through short activities like matching, reading, and listening exercises. Learners can earn points, keep streaks, unlock levels, and get instant

feedback based on their answers. Several studies have shown its benefits: (Payne, 2020) found that vocabulary knowledge improved significantly after four weeks of Duolingo use; (Grego, 2012) reported that 34 hours of using Duolingo was equal to one semester of college-level language learning; and Kazu and Kuvvetli (2025) said that the app's reward system motivates students to keep practicing. However, some studies have also reported weaknesses. Smiderle et al. (2020) warned that gamified learning can sometimes make students focus more on collecting rewards than understanding the material, which can lead to shallow learning and less motivation over time. Previous studies also differ in methods some are short-term experiments, others rely on self-reports and many have not tested Duolingo as part of formal classroom lessons, so their results may not apply to school settings.

In Indonesia, very few studies have examined the use of Duolingo in senior high school classrooms, especially for improving vocabulary needed in reading comprehension. Most research has looked at students using Duolingo on their own, not as part of guided classroom activities. This leaves a gap in understanding how gamified tools work when used alongside regular teaching in a formal school environment. The present study aims to fill that gap by testing the effectiveness of Duolingo as a vocabulary learning aid in senior high school EFL classes. This study aims to find out whether using Duolingo as part of classroom lessons can significantly improve students' vocabulary mastery. Specifically, it seeks to answer the question: Is there a significant difference in students' English vocabulary mastery before and after the implementation of Duolingo as a gamified learning tool?

Vocabulary Learning Theory

Nation (2022) states that vocabulary is essential for all four language skills, speaking, listening, reading, and writing and must be taught systematically. In the Indonesian EFL context, classroom time is often the only opportunity students have to develop their vocabulary, making it important to use methods that encourage active engagement and long-term retention. This theory supports the study's focus on measuring vocabulary mastery as the main outcome.

Vocabulary and Reading

Vocabulary knowledge directly affects reading comprehension (Aziez et al., 2020); (Abdullah Alzahrani et al., 2020). Students with a broader and deeper vocabulary can understand complex texts, make inferences, and learn from context. This link is relevant because the vocabulary taught through Duolingo is intended to help students understand the texts used in their reading lessons, making the app a tool to support classroom reading activities.

Gamification Theory

Gamification applies game design elements such as points, levels, and badges to learning activities to increase motivation and engagement (Deterding et al., 2011). Research shows that gamification can promote repeated practice, reduce anxiety, and support independent learning (Abin & Andas, 2022). In this study, gamification theory explains how Duolingo's reward system, streak tracking, and progress indicators could motivate students to practice vocabulary more regularly.

Duolingo in Gamified EFL Learning Environments

Duolingo combines vocabulary practice with gamification features such as XP points, streaks, levels, and instant feedback. These features are designed to encourage consistent practice and build vocabulary through repeated exposure (Rouabhia & Kheder, 2024). Teachers can also monitor students' progress through Duolingo's classroom dashboard,

making it easier to integrate into formal lessons. This theoretical basis justifies the use of Duolingo in this research as a supplementary tool for vocabulary learning.

METHOD

This study uses a quasi-experimental design. This design was chosen to assess how Duolingo's gamified learning has affected EFL students' reading skill and vocabulary mastery. Although participants in quasi-experimental research are not randomly assigned to groups as in true experimental designs, comparison and assessment of the treatment's efficacy are still (Creswell, 2018). In educational settings where full randomization is impractical because of administrative or ethical limitations, this type of design is commonly employed.

With reasonable internal validity, it allows researchers to investigate cause-and-effect relationships in authentic educational settings. This study was involved two group, one class as experiment group who received treatment and one class as control group who did not receive a treatment. Below is the tale of a nonrandomized two group design with pre-test and post-test. Group A was the Experimental group and Group B was the control group. Furthermore, y1 was the pre-test, y2 was the post-test and X was the Treatment.

Table 1. Two Groups

Group	Pre-Test	Independent Variable	Post-Test
A (Experimental)	y ₁	X	y ₂
B (Control)	y ₁	-	y ₂

Population and Sample of the Study

According to Johnson (2014) the population is the total set of people who have similar traits and to whom the results of the study will be applied. The population of this study is made up of all eight classes of eleventh-grade EFL students at a public high school in Trawas, Mojokerto. The sample was determined through purposive sampling, based on teacher recommendations.

On the advice of the English teacher, who found two eleventh-grade classes that were appropriate for the study, purposive sampling was used in this investigation. The teacher's suggestion was supported by a number of pedagogical factors, such as the students' regular attendance, involvement in English classes, and comfort with digital learning resources. Purposive sampling is appropriate when participants are chosen based on their potential to contribute rich and meaningful data as well as their relevance to the study's goals (Palinkas et al., 2015). In this regard, the chosen classes were thought to be perfect for gamified Duolingo activities because they had the qualities required to guarantee the success of the intervention. Following the teacher's well-informed advice, the researcher made sure the sample would be accessible and pedagogically appropriate for the study. Additionally, the sample included 62 students from two classes: XI 1 with 31 students as the control group, and XI 2 with 31 students as the experimental group.

Instruments

The main instrument in this research was a Vocabulary Mastery Test, used as both a pre-test and a post-test to measure students' vocabulary before and after the Duolingo-based learning activities. The test consisted of 25 multiple-choice questions designed according to CEFR A2–B1 standards. The questions were intended to measure both receptive vocabulary (recognizing words in reading) and productive vocabulary (using words appropriately). Identifying vocabulary meanings in context, identifying appropriate vocabulary forms within sentences, choosing words appropriate for given sentence

contexts, applying vocabulary in real-world everyday situations, and using vocabulary to understand and analyze short texts were the five categories into which the test items were divided. Al Kaboody (2023) on his review on Vocabulary in language teaching (2nd ed.), stated this structure expands on methodological suggestions for vocabulary assessment in EFL reading, and is consistent with Nation (2022) approach, which emphasizes form, meaning, and use.

The effectiveness of gamified vocabulary learning applications, especially Duolingo, in improving vocabulary acquisition among A2–B1 learners has also been confirmed by recent empirical studies by Mujiarni and Widiastuty (2024). The test problems cover a wide range of Bloom's revised taxonomy levels, from applying and analyzing to remembering and understanding, guaranteeing a thorough assessment of vocabulary knowledge in relation to the development of reading skills. Multiple choices question types were employed to provide consistency between the learning intervention and assessment. These question kinds closely resemble the task forms commonly found in the Duolingo program. The modification was made to maintain consistency between the instructional content provided through Duolingo and the skills assessed in the vocabulary tests, thereby increasing the validity and relevance of the test items in measuring students' vocabulary mastery. Both the pre-test and post-test were administered online via Google Forms to make the process more efficient and accessible for students. This method also ensured that all responses were automatically recorded and organized for analysis. The online delivery matched the digital nature of the Duolingo activities, creating consistency between the mode of learning and assessment.

Data Collection Technique

The data for this study was collected over four weeks during the second semester of the 2024–2025 school year. The process was designed to gather information about students' vocabulary mastery before and after using Duolingo in class. Before the study began, the researcher obtained permission from the school principal and the English teacher. Both teachers and students were informed about the objectives and procedures of the study, and consent was obtained. Ethical considerations such as voluntary participation and data confidentiality were maintained (Creswell, 2018). The research involved two groups: an experimental group that used Duolingo in addition to regular lessons, and a control group that only followed the regular lessons. Data was collected through pre-tests, post-tests, classroom observations, and Duolingo usage records. The particular steps taken at each meeting for both groups are listed in the following table:

Table 2. Data Collection Process

Meeting	Control Group	Experimental Group
Meeting 1 (Pre-Test)	- Students complete the Vocabulary Mastery Pre-Test to measure their initial vocabulary knowledge. - Students identify the structure and purpose of an Analytical Exposition Text through guided reading and discussion.	- Students complete the Vocabulary Mastery Pre-Test to measure their initial vocabulary knowledge. - Students identify the structure and purpose of an Analytical Exposition Text through guided reading and discussion. - Students recognize and operate the Duolingo application, including account creation and exploring key features.
Meeting 2 (Treatment)	- Students read and comprehend the content of an Analytical Exposition Text individually. - Students identify and note new vocabulary in the reading context and explain their meanings orally or in writing.	- Students use Duolingo for 15 minutes before class to practice and learn new vocabulary. - Students read and comprehend the content of an Analytical Exposition Text individually. - Students identify and note new

		vocabulary in the reading context and explain their meanings orally or in writing.
Meeting 3 (Treatment)	- Students write a short opinion paragraph using vocabulary they have learned from the text. - Students provide written peer feedback based on a simple rubric (structure, vocabulary usage, clarity of argument).	- Students use Duolingo for 15 minutes before class to practice and learn new vocabulary. - Students write a short opinion paragraph using vocabulary they have learned from the texts. - Students provide written peer feedback based on a simple rubric (structure, vocabulary usage, clarity of argument).
Meeting 4 (Post-Test)	To evaluate vocabulary acquisition progress, a Vocabulary Mastery Post-Test was given.	To evaluate vocabulary acquisition progress, a Vocabulary Mastery Post-Test was given.

Data Analysis

The data collected from the Vocabulary Mastery Tests (pre-test and post-test) were analyzed using quantitative statistical methods to determine the effect of gamified Duolingo activities on students' vocabulary mastery. The data analysis process included a number of statistical tests to guarantee the reliability and validity of the results:

Normality Test (Shapiro-Wilk)

It is crucial to confirm that the data distribution satisfies the assumption of normalcy prior to doing inferential statistical studies. The normality of the Vocabulary Mastery Test scores for the experimental and control groups in the pre-test and post-test was investigated in this study using the Shapiro-Wilk test. The normalcy threshold was set at a significance level of $\alpha = 0.05$. The data distribution can be regarded as normal if the p-value is higher than 0.05; if it is lower, the distribution is significantly out of the ordinary.

Homogeneity Test

Examining the homogeneity of variance between the experimental and control groups came next in the analysis once the data's normality was confirmed. When doing parametric tests like the independent samples t-test, it is crucial to assume that the variances of the two groups are equal, which is known as homogeneity of variance (Ary et al., 2010). The homogeneity of the Vocabulary Mastery Test results from the pre-test and post-test was evaluated in this study using Levene's Test for Equality of Variances. For Levene's test, the null hypothesis (H_0) asserts that the group variances are equal. To ascertain the outcome, a significance level of $\alpha = 0.05$ was employed. The variances are homogeneous if the p-value is larger than 0.05; heterogeneity is suggested if it is smaller.

Initial Test

To determine if the experimental and control groups' levels of vocabulary competency were comparable, an initial test was given prior to the commencement of the treatment. To ensure that any differences observed in post-test results are caused by the therapy and not by inherent group differences, this step is essential (Creswell, 2018). An Independent Samples t-test was used to compare the mean pre-test scores of the two groups. According to the null hypothesis (H_0), the experimental and control groups' pretreatment vocabulary mastery did not differ significantly.

Within-Group Test

A within-group study was carried out by comparing students' pre-test and post-test results using a Paired Samples t-test in order to ascertain the efficacy of training within each group. The significance of change within the same group over two time points can be measured

with this statistical test (Creswell, 2018). The experimental and control groups were subjected to separate applications of the Paired Samples t-test. The test sought to determine whether using gamified Duolingo tasks significantly improved vocabulary mastery in the experimental group. The same examination looked at the effects of conventional vocabulary training in the control group.

Hypothesis Testing (Independent Sample t-test)

To determine whether there was a statistically significant difference in vocabulary knowledge between the experimental and control groups after the therapy, an Independent Samples t-test was employed. This test can be used to assess if gamified Duolingo activities produced a noticeable improvement over conventional learning when comparing the means of two distinct groups (Creswell, 2018). The significance level was set at $\alpha = 0.05$. The Levene's Test for Homogeneity had previously confirmed that the test was done under the presumption of equal variances. The post-test findings from both groups were then entered into statistical software (like SPSS).

RESULTS

Descriptive Statistics

This section presents descriptive statistics of students' vocabulary mastery on reading skill outcome in the experimental and control groups. These data, which give a preliminary picture of the students' abilities before and after the treatment, include the mean scores of the pre-test and post-test.

Table 3. Descriptive Statistics

Group	Test Type	N	Mean	Std. Deviation
Experimental	Pre-test	31	72.06	12.312
Experimental	Post-test	31	82.00	12.170
Control	Pre-test	31	66.14	15.103
Control	Post-test	31	74.93	11.584

The mean score of the experimental group was 72.06 with a standard deviation of 12.312 before the exam, and it rose to 82.00 with a somewhat lower standard deviation of 12.170 after the test. This suggests that the intervention (gamified Duolingo activities) was effective because it shows a discernible improvement in the students' reading skills after receiving the therapy. In contrast, the control group's mean scores were 66.14 on the pre-test and 74.93 on the post-test, with standard deviations of 15.103 and 11.584 for each group, respectively. Even though the control group also saw improvement, it was comparatively less than that of the experimental group. Although both groups generally showed progress, the experimental group's average increase was larger, suggesting that the intervention may have had a greater favorable impact on their reading achievement.

Test of Normality

It is crucial to determine whether the data are normally distributed in order to guarantee the suitability of parametric statistical analyses, such the t-test. The normality of the pre-test and post-test results for the experimental and control groups was evaluated in this study using the Shapiro-Wilk test. The Shapiro-Wilk test assesses the null hypothesis that the data have a normal distribution and is regarded as reliable, particularly for small to moderate sample sizes ($n < 50$).

Table 4. Normality Test

Group	Test Type	Statistic	df	Sig. (p-value)
Experimental	Pre-test	0.951	31	0.121
Experimental	Post-test	0.943	31	0.072
Control	Pre-test	0.963	31	0.282
Control	Post-test	0.971	31	0.469

The data in all four groups do not statistically deviate from a normal distribution, according to the above table, as all p-values (Sig.) are higher than the significance level of 0.05. To be more precise, the experimental group's post-test score has a p-value of 0.072 while its pre-test score has a p-value of 0.121. The pre-test and post-test p-values for the control group are 0.282 and 0.469, respectively. The Shapiro-Wilk test's null hypothesis, according to which the data are regularly distributed, is accepted since $p > 0.05$ in every instance.

This enables the use of parametric tests, like the independent samples t-test and paired samples t-test, in further analyses by confirming that the assumption of normalcy is satisfied. To sum up, the results of the normality test confirm that using parametric inferential approaches to investigate the variations in reading proficiency between the experimental and control groups is valid.

Test of Homogeneity

Examining whether group variances are equal is a prerequisite for performing parametric tests like the independent samples t-test. Levene's examine is used to examine this assumption, which is referred to as the homogeneity of variance. The null hypothesis, according to which the variances of the two groups are equal, is assessed by the Levene's Test. The assumption of equal variances is deemed satisfied if the significance value (Sig.) is higher than 0.05.

Table 5. Test of Homogeneity of Variance

Test Type	Levene's Statistic	df1	df2	Sig. (p-value)
Post-test	1.274	1	60	0.264
Post-test	0.248	1	60	0.620

The p-values for the pre-test (0.264) and post-test (0.620), as indicated in the table, are both higher than the alpha threshold of 0.05. This suggests that the experimental and control groups' variances on the pre-test and post-test scores do not differ significantly. The p-value for the pre-test is 0.264, and the Levene Statistic is 1.274. The p-value for the post-test is 0.620, and the Levene Statistic is 0.248.

The null hypothesis of equal variances is accepted because neither of the p-values is significant ($p > 0.05$). This indicates that the homogeneity of variance assumption is met, therefore comparing the means of the two groups using independent sample t-tests is statistically justified. Levene's Test results verify that the data satisfy the homogeneity of variance assumption. By doing this, the validity of later parametric inferential studies is strengthened.

Test of Initial Equality

An independent samples t-test was used to examine the pre-test results in order to ascertain if the experimental and control groups started with similar reading proficiency levels before the treatment. The purpose of this test is to assess the two groups' initial equality prior to the application of the intervention (gamified Duolingo tasks). The mean pre-test scores of the experimental and control groups do not differ significantly, according to the null

hypothesis (H_0) for this study. If this hypothesis were to be rejected, it would imply that the groups were not initially equal, which might jeopardize the reliability of any further comparisons.

Table 6. Initial Equality Test

N	Group	Mean	Standard Deviation	t	df	Sig. (2-tailed)
31	Experimental	72.06	12.312	1.733	60	0.085
31	Control	66.14	15.103			

The experimental group's mean pre-test score was 72.06 (SD = 12.312), but the control groups was 66.14 (SD = 15.103), as indicated in Table 4.4. With 60 degrees of freedom, the independent samples t-test produced a p-value of 0.085 and a t-value of 1.753. The outcome is not statistically significant since the p-value exceeds the significance level of 0.05.

This suggests that the experimental and control groups' pretreatment starting reading proficiency did not differ significantly. Thus, it may be said that the two groups' starting reading proficiency levels were equal, and the null hypothesis (H_0) is accepted. The hypothesis that the experimental and control groups began from a similar baseline is supported by the results of the test of initial equality. This increases the study's design's credibility by guaranteeing that any variations in post-test scores that are noticed are due to the treatment and not to underlying inequalities.

Within-Group Improvement

A descriptive comparison of the mean pre-test and post-test scores was carried out in order to evaluate the efficacy of the treatment within each group. Finding out how much the experimental and control groups' kids' reading performance improved before and after the intervention is the main goal of this investigation. The gain score, which represents the learning progress made individually by each group, is the difference between the post-test and pre-test results.

Table 7. Paired t-Test Results: Pre-Test vs. Post Test in Each Group

Group	Test Type	Mean	Gain (Post – Pre)
Experimental	Pre-test	72.06	9.94
Experimental	Post-test	82.00	
Control	Pre-test	66.14	8.79
Control	Post-test	74.93	

Both groups saw increases in their reading scores from the pre-test to the post-test, as the table illustrates. After participating in gamified learning activities with Duolingo, the experimental group's mean score increased from 72.06 on the pre-test to 82.00 on the post-test, yielding a gain of 9.94 points. This suggests that reading abilities significantly improved after the intervention. From a pre-test mean of 66.14 to a post-test mean of 74.93, the control group, who received traditional training, also showed an increase of 8.79 points. Even while there was an improvement, the gain was not as large as it was for the experimental group.

The experimental group showed a higher average gain, which may indicate that the gamified approach had a bigger influence on improving EFL students' reading performance, even if these results show that both teaching approaches helped students increase their reading skills. Students in both the experimental and control groups benefited from their respective teaching strategies, according to the within-group gain analysis. The

experimental group's greater growth, however, offers early proof that gamified Duolingo exercises are more successful than conventional approaches at enhancing students' reading abilities.

Hypothesis Testing (Independent sample t-test)

An independent samples t-test was used to assess whether the experimental and control groups' post-intervention reading skill results differed statistically significantly. As demonstrated in the preceding sections (4 and 5), this test is applicable when comparing the means of two independent groups and the assumptions of normality and homogeneity of variance are satisfied.

According to the null hypothesis (H_0), the experimental and control groups' post-test mean scores do not differ significantly. The alternative hypothesis (H_1), on the other hand, contends that there is a substantial difference in the post-test mean scores, demonstrating the impact of the intervention (in this study, gamified Duolingo activities).

Table 8. Independent t-Test of Post-test Scores

Group	N	Mean	SD	t	df	Sig. (2-tailed)
Experimental	31	85.00	10.500	4.205	60	<0.001
Control	31	72.00	11.200			

The experimental group's post-test mean score was 85.00 (SD = 10.50), whereas the control groups was 72.00 (SD = 11.20). With 60 degrees of freedom, the independent samples t-test yielded a t-value of 4.205 and a matching p-value of less than 0.001. The null hypothesis is rejected since the p-value is less than the significance level ($\alpha = 0.05$). This suggests that there is a statistically significant difference between the experimental and control groups' post-test results. To put it another way, the experimental group's increase is not the result of chance; rather, it shows how well the intervention gamified Duolingo activities worked to improve the reading abilities of EFL students.

The experimental group fared better on the post-test than the control group, according to the results of the hypothesis test. The conclusion that gamified learning tactics, especially those used with Duolingo, are more effective than traditional approaches in enhancing students' reading skill in an EFL context is supported by the considerable difference in mean scores.

DISCUSSION

The findings of this study provide strong evidence that using Duolingo as a gamified learning tool can significantly improve EFL students' vocabulary mastery and reading comprehension. This result is consistent with previous studies, such as Abin and Andas (2022), who found that Duolingo improved vocabulary development among junior high school students. Similarly, Kazu and Kuvvetli (2025) reported that gamification features like XP points, streaks, and instant feedback increased motivation and practice frequency, both of which are essential for vocabulary growth. In this study, the experimental group showed a statistically significant improvement in post-test scores compared to the control group. This supports Nation (2022) claim that vocabulary mastery requires repeated and meaningful exposure to target words. Duolingo's use of spaced repetition and contextualized exercises appears to provide this type of exposure, which may explain the students' progress.

Beyond vocabulary, the results also showed improvements in reading comprehension. This aligns with Abdullah Alzahrani et al. (2020) who emphasized that vocabulary size is one of the strongest predictors of reading comprehension, especially in

EFL contexts where students rely heavily on written texts. The experimental group demonstrated better understanding of texts, greater confidence in reading, and improved ability to apply new vocabulary in comprehension tasks. These outcomes are also in line with Rouabhia and Kheder (2024) who reported similar gains from integrating Duolingo into formal lessons. The results also highlight an important shift from traditional vocabulary instruction. Unlike rote memorization of word lists, Duolingo provides vocabulary practice within interactive, meaningful contexts. This supports Al-Dosakee and Ozdamli (2021) conclusion that gamification can enhance learner motivation and performance. The combination of engaging activities and immediate feedback seems to encourage sustained learning, while the relaxed, game-like environment may reduce learning anxiety. This was evident in classroom observations, where students reported feeling more comfortable experimenting with vocabulary and making mistakes an important condition for language development, as noted by Jane et al. (2024)

However, the findings also point to potential limitations of gamification. As Smiderle et al. (2020) observed, gamified tools may sometimes lead to superficial engagement, with learners focusing more on earning rewards than understanding content. Similar behavior was noted in this study, as some students were more concerned with maintaining their Duolingo streaks than fully engaging with the material. This suggests that while gamification can be effective, it must be carefully integrated into lessons with teacher guidance to ensure that learning goals remain the focus. Overall, the results support the initial hypothesis that Duolingo would have a significant positive effect on vocabulary mastery and reading comprehension. At the same time, they show that the impact of gamified tools may vary depending on learners' motivation levels, learning styles, and willingness to engage meaningfully with the content. Alternative interpretations of the results might consider the role of teacher support, peer interaction, or even students' prior familiarity with digital learning tools in influencing the outcomes.

To sum up, these findings reinforce the idea that blended learning approaches combining traditional instruction with interactive digital tools can be highly effective in EFL classrooms. By integrating Duolingo into structured, teacher-led lessons, educators can make vocabulary learning more engaging and relevant, while also supporting the development of reading skills. This approach not only meets curriculum goals but also promotes learner autonomy and motivation, ultimately contributing to better language learning outcomes.

CONCLUSION

The results of this study clearly show that the use of Duolingo as a gamified learning tool significantly improved the vocabulary mastery of EFL senior high school students. The experimental group, which received Duolingo-based vocabulary activities alongside their regular lessons, achieved a statistically significant gain in vocabulary scores compared to the control group ($p < 0.001$). This improvement was reflected not only in the inferential statistics but also in the descriptive data, which showed a higher mean gain for the experimental group.

This finding directly answers the research question: Is there a significant difference in students' English vocabulary mastery before and after the implementation of Duolingo as a gamified learning tool? The evidence from this study confirms that such a difference exists and that the improvement is both meaningful and measurable.

Several factors appear to have contributed to this result. Duolingo's gamification features, such as XP points, levels, streaks, and instant feedback, created a motivating and low-anxiety environment that encouraged consistent practice. Its spaced repetition system and contextualized vocabulary exercises supported better vocabulary recall and application, aligning with Nation's (2022) theory that repeated and meaningful exposure is essential for mastery. Furthermore, integrating Duolingo with classroom reading activities

provided students with opportunities to apply new vocabulary in meaningful contexts, strengthening both receptive and productive vocabulary skills.

In summary, this study supports the integration of gamified tools like Duolingo into EFL classrooms as part of a blended learning strategy. By combining structured teacher-led instruction with interactive digital activities, educators can make vocabulary learning more engaging, promote learner autonomy, and improve reading comprehension alongside vocabulary development. The positive outcomes observed here suggest that Duolingo is not only an effective supplementary tool for vocabulary learning but also a means to create a more student-centered and motivating language learning environment.

REFERENCES

- Abdullah Alzahrani, S., Fung Yong, M., Jalaluddin, I., Md Rashid, S., & Mei Fung, Y. (2020). Relationship Between Vocabulary Size and Reading Comprehension Achievement Among Saudi High School EFL Learners. In *Hamdard Islamicus* (Vol. 43, Issue 2). <https://www.researchgate.net/publication/367559556>
- Abin, R., & Andas, N. H. (2022). The Effect of Using Autonomous Learning Model by Duolingo-Assisted to Improve Students' Vocabulary Mastery at VIII Class of SMPN 3 Buton Tengah. *EduLine: Journal of Education and Learning Innovation*, 2(4), 415–426. <https://doi.org/10.35877/454ri.eduline1258>
- Al Kaboody, M. (2023). (2023). *Review of Schmitt, N., & Schmitt, D. (2020). Vocabulary in language teaching (2nd ed.)*.
- Al-Dosakee, K., & Ozdamli, F. (2021). Gamification in Teaching and Learning Languages: A Systematic Literature Review. *Revista Romaneasca Pentru Educatie Multidimensionala*, 13(2), 559–577. <https://doi.org/10.18662/rrem/13.2/436>
- Ary, D., Jacobs, L. C., & Sorensen, C. (2010). *Introduction to Research in Education*.
- Aziez, F., Aziez, F., & Purwanto, B. E. (2020). Receptive vocabulary knowledge and reading skills of Indonesian prospective EFL teachers. *Universal Journal of Educational Research*, 8(5), 2005–2011. <https://doi.org/10.13189/ujer.2020.080538>
- Creswell, J. W. , & C. J. D. (2018). *Fifth Edition Research Design Qualitative, Quantitative, and Mixed Methods Approaches*.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining “gamification.” *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*, 9–15. <https://doi.org/10.1145/2181037.2181040>
- Grego, J. (2012). *Duolingo Effectiveness Study Final Report Executive Summary*.
- Jane, O. C., Ezeonwumelu, C. G., Barah, C. I., & Jovita, U. N. (2024). The Role of Gamification in Enhancing Foreign Language Education: Strategies, Benefits, and Case Studies. *Newport International Journal of Research in Education*, 4(1), 33–38. <https://doi.org/10.59298/NIJRE/2024/41133388>
- Johnson, R. B. , & C. L. B. (2014). *Educational Research Quantitative, Qualitative, and Mixed Approaches Fifth Edition*.
- Kazu, İ. Y., & Kuvvetli, M. (2025). Digital language learning with Duolingo: assessing its impact on listening, speaking, reading, and writing skills. *Journal of Computers in Education*. <https://doi.org/10.1007/s40692-025-00355-0>
- Mujiarni, U., & Widiastuty, H. (2024). *Duolingo for Vocabulary Mastery: Empowering Students in the Digital Age of Language Learning* (Vol. 8, Issue 2).
- Nation, I. S. P. (2022). *Learning Vocabulary in Another Language 2nd Edition*. <https://ebookgate.com/product/in-another-country-priya-joshi/>
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and*

- Mental Health Services Research*, 42(5), 533–544. <https://doi.org/10.1007/s10488-013-0528-y>
- Payne, J. S. (2020). Developing L2 productive language skills online and the strategic use of instructional tools. *Foreign Language Annals*, 53(2), 243–249. <https://doi.org/10.1111/flan.12457>
- Rouabhia, R., & Kheder, K. (2024). Using Duolingo in Teaching and Learning Vocabulary: A Systematic Review. *Indonesian Journal of English Language Studies (IJELS)*, 10(2), 91–107. <https://doi.org/10.24071/ijels.v10i2.8873>
- Smiderle, R., Rigo, S. J., Marques, L. B., Peçanha de Miranda Coelho, J. A., & Jaques, P. A. (2020). The impact of gamification on students' learning, engagement and behavior based on their personality traits. *Smart Learning Environments*, 7(1). <https://doi.org/10.1186/s40561-019-0098-x>