

Enhancing English Achievement through a Duolingo-Based Learning Intervention: The Role of Usage Intensity in Learning Gains

Yulia Wahyuningsih¹ Universitas PGRI Kanjuruhan Malang, Indonesia¹
yuli79w@gmail.com¹

Mujiono² (✉) Universitas PGRI Kanjuruhan Malang, Indonesia²
moejie_nova@unikama.ac.id²

Lasim Muzammil³ Universitas PGRI Kanjuruhan Malang, Indonesia³
muzammil_lasim@unikama.ac.id³

 <https://doi.org/10.58194/eloquence.v5i1.3265>

Corresponding Author: ✉ Mujiono

Article History

Received
11-12-2025
Accepted:
31-01-2026
Published:
04-04-2026

ABSTRACT

Background: The rapid growth of MALL technology has positioned Duolingo as one of the top language learning platforms in the modern era. However, uncertainty persists regarding how variable user frequency affects outcomes, notably when learners have diverse initial skill levels.

Purpose: The study examined whether an individual's initial level of performance influenced a structured Duolingo activity that improved their English skills. The study further explored the correlation between the application of intensive treatment and outcomes after testing.

Method: This study used a one-group pretest–posttest quasi-experimental design. It involved 60 EFL learners completing a structured Duolingo intervention. Standardized tests were used to measure proficiency before and after the program. Usage logs record session timing and duration. ANCOVA was used to test the effectiveness of the intervention and the predictive role of usage intensity.

Results and Discussion: Descriptively, the low-intensity group achieved the highest unadjusted posttest mean ($M = 80.40$). However, interpretation relied on ANCOVA-adjusted posttest scores controlling for pretest performance. The results showed a significant effect of usage intensity, $F(2,57) = 4.421$, $p = .016$, $\eta^2 = .128$. Learners with moderate-to-low engagement demonstrated greater learning gains than high-intensity users, indicating that regulated use supports more effective English learning.

Conclusions and Implications: Results showed that learning English through Duolingo was effective when users were actively engaged and provided with appropriate guidance. This brings the study's full circle to the point that what matters most is the fairness of involvement. Therefore, it is recommended that protocols be established to prevent excessive use of this platform. These strategies must be sufficiently data-informed to preserve effectiveness, sustainability, and pedagogical soundness of technologically equipped language instruction.

Keywords:

Duolingo; Mobile-Assisted Language Learning; Usage Intensity; English Proficiency; Quasi-Experimental Design.

ABSTRAK

Latar Belakang: Perkembangan pesat teknologi MALL telah menjadikan Duolingo salah satu platform pembelajaran bahasa terkemuka. Namun, masih terdapat ketidakpastian mengenai pengaruh variasi frekuensi penggunaan terhadap hasil belajar, terutama pada peserta didik dengan kemampuan awal yang berbeda.

Tujuan: Penelitian ini menganalisis pengaruh tingkat kemampuan awal terhadap efektivitas aktivitas Duolingo terstruktur dalam meningkatkan keterampilan bahasa Inggris, serta hubungan antara perlakuan intensif dan hasil pascates.

Metode: Penelitian ini menggunakan desain quasi-experimental dengan satu kelompok pretest–posttest. Sebanyak 60 pembelajar EFL mengikuti intervensi terstruktur melalui Duolingo. Tes standar menilai kemahiran sebelum dan sesudah program, sedangkan log penggunaan mencatat waktu serta durasi sesi. Analisis ANCOVA digunakan untuk mengukur efektivitas intervensi dan peran prediktif intensitas penggunaan.

Hasil dan Pembahasan: Secara deskriptif, kelompok dengan intensitas penggunaan rendah mencapai rerata posttest tidak disesuaikan tertinggi ($M = 80,40$). Namun demikian, interpretasi didasarkan pada skor posttest yang telah disesuaikan melalui ANCOVA dengan mengontrol kinerja pretest. Hasil analisis menunjukkan adanya pengaruh signifikan dari intensitas penggunaan, $F(2,57) = 4,421$, $p = 0,016$, $\eta^2 = 0,128$. Peserta didik dengan keterlibatan moderat hingga rendah menunjukkan peningkatan hasil belajar yang lebih besar dibandingkan pengguna dengan intensitas tinggi, yang mengindikasikan bahwa penggunaan yang terregulasi mendukung pembelajaran bahasa Inggris yang lebih efektif.

Kesimpulan dan Implikasi: Hasil menunjukkan bahwa pembelajaran bahasa Inggris melalui Duolingo efektif ketika pengguna terlibat aktif dan mendapat bimbingan yang sesuai. Ini menegaskan bahwa keterlibatan yang adil lebih penting bagi keberhasilan pembelajaran. Oleh karena itu, disarankan untuk menetapkan protokol guna mencegah penggunaan berlebihan platform ini. Strategi tersebut harus berbasis data untuk menjaga efektivitas, keberlanjutan, dan ketepatan pedagogis pembelajaran bahasa berbasis teknologi.

Kata Kunci

Duolingo; Pembelajaran Bahasa Berbantuan Mobile (MALL); Intensitas Penggunaan; Kemahiran Bahasa Inggris; Desain Kuasi-Eksperimental.



Copyright: © 2026 by the author(s).

This is open access article under the

[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

INTRODUCTION

The growing trend of worldwide globalization and digital revolution has made English fluency as an essential aspect, both from a linguistic and sociopolitical perspective. English is no longer restricted to being a language in foreign countries; it has now become a universal literacy, the foundation for participating in various sectors of education, professions, and communication across nations.[1], [2] This universal need for English proficiency worldwide has led to the integration of technology into language instruction by governments and institutions, enabling them to cater to the learning needs of large numbers of people. However, in many English as a Foreign Language (EFL) contexts, performers fail to deliver their best possible performance due to several reasons, including insufficient classroom exposure, limited authentic interaction, and reduced learning autonomy.[3], [4] This has been the primary reason why technology-integrated language learning, particularly in terms of Mobile-Assisted Language Learning (MALL), has been drawing comparatively more attention because it is classified as flexible, scalable, and learner-centric.[5], [6]

As far as MALL platforms are concerned, Duolingo has been at the top of the list due to its influence on the public and the examination of the application by multiple researchers. Primarily, it was the combination of multiparticipant aspects, constructed feedback, and split-second learning strategies that became the first line of personalization for Duolingo.[3], [7] The results of several studies have revealed that mobile gamification leads to and encourages persistence, motivation, and

self-directed learning behavior—three key factors in efficient and successful second language acquisition.[6], [8] A significant aspect of the Duolingo application is its reward-based mechanism, which helps sustain attention and alleviate learning anxiety, as well as its data-driven adaptivity feature that provides support for progress tracking per student across different skill levels simultaneously [4]. Furthermore, meta-analyses have demonstrated that MALL-based interventions have a moderate to strong positive influence on English achievement compared to traditional classroom instruction.[1], [5] Therefore, it is suggested that the investigated application could be a structured pedagogical intervention, rather than a first-class educational technology alone.

However, despite the bright prospects, one significant concern is looming: how the intensity metrics of Duolingo usage—number of sessions, session length, and cumulative time—contribute to learning outcomes. The primary focus of past studies has been on the adoption of the app and the satisfaction of app users, rather than the quantitative behavioral metrics that can predict language attainments [9], [10]. [2] have emphasized that research into MALL still primarily follows a pattern of quantitative methods and suggested the necessity of designs that would take a long time to observe users' behaviors mentally. Along the same lines of thought, [11] found that Moodle-based tasks helped students develop more fluent writing and lecture-taking skills; however, at the time, they did not control for the number of hours students spent online. This flaw renders the question of how the activities students are involved in impact their learning outcomes an acidic topic ([5], [8], [11]). With the proliferation of digital learning, determining usage intensity as a predictor is essential for fine-tuning technology-mediated language learning theories.

The focus of the current research is primarily on language skills at the lower levels, rather than broader measures of total English proficiency that are often overlooked in the process.[12], [13] In other words, one of the examples of this is that Duolingo is a platform that can improve vocabulary memorization as well as the learner's motivation nonetheless the relationship of this improvement with the achievement of the skills of reading, writing, and speaking in the long run is still not very clear .[14] The lack of commensurable data for different skills developed during the study is a significant problem, as it limits not only the validity of the findings but also their application to curriculum integration. Furthermore, a significant portion of the interventions, rather than occurring within a formal setting, take place outside the walls of schools and therefore do not utilize the advantages of systematic assessment.[2] Hence, it is not yet known to what degree the use of mobile learning apps serves as a supplement to the traditional classroom or can even be the primary predictor of full language proficiency.[5], [7]

A significant amount of research has demonstrated that the use of MALL is also linked to emotional and motivational dimensions . Learners' opinions about the app's utility and their state of being in flow significantly impact their motivation and persistence in learning.[4] Similarly, [9] demonstrated through their study that collaborative and gamified tasks in mobile learning contribute to the development of attention and self-regulation—two essential elements for enhancing learning intensity . At the same time, the interplay between high usage of MALL and student achievement included not just the traditional view of student time allocation (behavioral engagement) but also psychological engagement (motivation, efficacy, and persistence) . In addition, studies carried out by [1], and [5] confirmed that the regularity of learner activity significantly contributes to the effectiveness of MALL interventions. In other words, the very core of enhancing the digital language learning field is to gain insight into the interplay of technology use, motivation, and excellence.[8], [10]

A pedagogical point of view highly values the identification of usage intensity as a predictive factor, as it serves as a linkage between the technological sector and learning analytics. Although general learning platforms like Duolingo, which are app-based, gather a significant amount of data from users, little to no is known about the reliability of this data and its educational relevance in

empirical research.[3], [4] The determination of engagement would be instrumental in making data-driven decisions regarding teaching methods, the delivery of adaptive content, and the design of an evidence-based curriculum.[2] Furthermore, by quantifying the amount of time learners spend interacting with an app, it may be possible to determine the limits beyond which the degree of engagement is deemed significant enough to warrant a corresponding increase in learning rates. This study addresses a gap in the MALL literature concerning the effects of usage intensity. It examines frequency and duration within a Duolingo-based intervention using an ANCOVA-controlled design. The pretest–posttest framework controls for initial proficiency to estimate adjusted learning gains accurately. By moving beyond adoption-focused studies, it provides novel, analytics-informed evidence for personalized mobile language learning. Accordingly, this study aims to examine the effectiveness of a Duolingo-based intervention and the predictive role of usage intensity in shaping students' English achievement across skill domains. The first objective is to evaluate whether structured integration of Duolingo significantly improves English proficiency compared with conventional learning. The second is to assess whether usage intensity metrics—session frequency, duration, and total engagement time—predict learners' overall achievement levels. This dual approach allows for testing both the causal impact and the predictive value of app usage behaviours.[3], [5]

Theoretical implications of the study go beyond the current MALL frameworks by suggesting that usage intensity is a quantifiable and explanatory factor in digital language learning models.[4], [10] Practically, this implies that educators and policymakers can apply this research to their work in designing, implementing, and monitoring app-based learning interventions.[1], [2] Moreover, one of the main findings of the study is that app developers can utilize it as a source of information to fine-tune adaptive algorithms, which will not only attract users but also foster deep learning. The study ultimately relates to the broader issue of the role of technology, motivation, and frequency of use in contributing to progress in English language learning worldwide. The problem addressed in this study is formulated as follows: To what extent does the Duolingo-based learning intervention enhance students' English achievement after adjusting for initial proficiency (pre-test), and to what extent does Usage Intensity significantly predict students' post-intervention English achievement once initial proficiency is statistically accounted for?

LITERATURE REVIEW

1. Duolingo as a Mobile-Assisted Language Learning Platform

Duolingo has emerged as one of the most widely used MALL apps globally, offering engaging and playful micro-lessons that emphasize the importance of repetition, feedback, and gamification. Another remarkable aspect of it is that it addresses learners' needs by utilizing a system that aligns with the communicative and cognitive theories of second language learning. As a consequence, users can get involved through every day, short, yet high-quality learning tasks that lead to the accumulation of skills and knowledge.[6], [15] The question about the value of leveraging the app for learning purposes has been raised and is driving researchers to explore the matter and carry out studies in both school and non-school contexts.

Using Duolingo in a structured manner can promote speaking and vocabulary retention among EFL learners [6]. [15], also agreed that this digital learning system brought a noticeable enhancement in the reading and listening skills of the frequent users. The evidence from the two studies thus shows that the primary strength of Duolingo is in bringing continuous engagement. However, the success of Duolingo is not solely reliant on its structure, but also on learners' actions and attitudes toward the application over a long period, thus emphasizing the necessity of measuring Duolingo Usage Intensity (DUI) as a direct determinant of learning outcomes. DUI and Learning Outcomes

The concept of Duolingo Usage Intensity (DUI) is a basic measure that covers the aspects of frequency, period, and regularity of learners using Duolingo's instructional content. [16] have found a high correlation between the DUI measure and the remarkable progress in learners' vocabulary and grammar accuracy, indicating that moving forward will be largely dependent on the intensity the user chooses. On the other hand, [17] have calculated that sacrificing 34 hours of active Duolingo use is equivalent to taking a semester of language instruction, revealing a correlation between students' backgrounds and app usage for language learning.

Recent studies have shown that students who consistently practice and learn for extended hours daily tend to achieve the top marks in proficiency tests.[18] However, the impact of DOI might hinge on the user's motivations and relations to the system: learners with a high level of self-regulation are more likely to be consistent in using the app compared to passive users.[13], [19] The findings in this field describe DUI as a sign of both behavioral and emotional aspects of digital learning success.

2. Gamification in Duolingo: Motivation and Sustained Learner Engagement

At the centre of Duolingo's engagement strategy is the use of gamification. Criteria such as XP points, streaks, leaderboards, and virtual rewards motivate people to persist by providing immediate reinforcement.[19] These elements of design are found to align with self-determination theory, providing users with the competency and social relation—two needs that sustain motivation for the learning process. [13] noted that play engagement in Duolingo enhances learners' feelings of success and promotes consistent activity, thanks to the same gamification mechanics.

Nonetheless, once extrinsic rewards become more important than intrinsic motivation, the students' commitment may be superficial—they will log in to maintain their streaks without genuinely engaging mentally with the exercises.[6], [16] For any effective DUI, gamified activities should strike a balance between motivation and engagement with the content. Therefore, we conclude that an understanding of the role of motivation in driving user success is essential for predicting the success of Duolingo users.

METHOD

1. Research Design

This study employed a quasi-experimental one-group pretest–posttest design to examine the effect of a Duolingo-based learning intervention on students' English proficiency. In this design, all participants received the same instructional treatment, and learning outcomes were evaluated by comparing pretest and posttest scores within the same group. This approach enabled the examination of learning gains attributable to the intervention in an authentic educational setting where random assignment and control groups were impractical.[20], [21]

Within this one-group framework, usage intensity was treated as a grouping variable rather than an experimental manipulation. Usage intensity was operationalized based on learners' behavioral engagement during the intervention and subsequently categorized into three levels (low, moderate, and high) to capture natural variation in learning behavior. This analytical strategy is consistent with prior quasi-experimental studies that investigate differential outcomes within a single cohort using engagement-based classifications.[22], [23]

To account for differences in learners' initial English proficiency, Analysis of Covariance (ANCOVA) was employed. In this analysis, pretest scores served as the covariate, posttest scores as the dependent variable, and usage intensity level (low, moderate, high) as the grouping factor. ANCOVA was selected because it allows posttest comparisons while statistically controlling for baseline differences, thereby reducing bias and improving the accuracy of effect estimation in non-randomized designs.[22], [23] Such an approach is particularly suitable for technology-enhanced language learning research, where behavioral engagement varies naturally among learners and constitutes an important explanatory factor.[24]

2. Participants

This study involved 60 junior high school students aged 14–15 years. All participants were enrolled in English classes in which Duolingo was integrated as an instructional support tool. Initial English proficiency was established through a standardized pretest. A total sampling technique was applied due to the homogeneous and limited population. Inclusion criteria required continuous enrollment, completion of both pretest and posttest, and consistent Duolingo use verified by usage logs; students with incomplete data were excluded. The analysis focused on intra-group variability in engagement and learning gains, consistent with prior quasi-experimental research.[25]; [26]

3. Duolingo-Based Intervention and Usage Intensity

The intervention was conducted over a six-week period, targeting vocabulary, grammar, and reading comprehension through self-directed Duolingo tasks aligned with course objectives. Appropriate guidance was provided via weekly learning targets, progress monitoring, and instructor feedback to support meaningful engagement. Usage intensity was defined using total learning time, number of sessions, and accumulated experience points (XP) extracted from usage logs. A composite score was calculated, and participants were classified into low-, moderate-, and high-intensity groups using a tertile-based cut-off, enabling comparison of learning outcomes across engagement levels.[25]; [26]

4. Instruments

This study employed three main instruments: an English proficiency test administered as both a pretest and posttest, and a Duolingo usage log. The English proficiency test was a researcher-developed standardized instrument aligned with the Common European Framework of Reference (CEFR) levels A1–B1, designed to measure learners' vocabulary, grammar, and reading comprehension.

Table 1. CEFR-Aligned Question Blueprint

CEFR Level	Skill Focus	Subskill Assessed	Cognitive Process	Number of Items	Illustrative Descriptor
A1	Reading Comprehension	Identifying explicit information	Retrieval / Recognition	8	Understands familiar words and simple sentences
A1	Reading Comprehension	Understanding basic	Basic Interpretation	6	Recognises everyday

		vocabulary in context			expressions and phrases
A2	Reading Comprehension	Identifying main ideas	Interpretation	10	Understands short, simple texts on familiar matters
A2	Reading Comprehension	Understanding relationships between ideas	Relational Processing	8	Extracts meaning from connected sentences
B1	Reading Comprehension	Inferencing meaning from context	Higher-Order Interpretation	10	Understands texts describing events and experiences
B1	Reading Comprehension	Interpreting implicit information	Analytical Processing	8	Identifies viewpoints and implied meanings

A test blueprint guided item development to ensure proportional representation of language skills and difficulty levels. Sample indicators included identifying word meanings in context (vocabulary), selecting grammatically appropriate sentence structures (grammar), and answering factual and inferential questions based on short reading passages (reading comprehension). All items were multiple-choice and constructed following established principles of language assessment, emphasizing content relevance and alignment with instructional objectives.

The Duolingo usage log functioned as a behavioral measurement instrument and automatically recorded total learning time (minutes), number of learning sessions, completed units, and accumulated experience points (XP). These behavioral indicators were used to quantify learners' engagement during the intervention and served as the basis for calculating a Usage Intensity Score. Digital learning logs have been widely recognized as reliable indicators of behavioral engagement in gamified language learning environments.[23], [24]

5. Validity and Reliability of the Instruments

Content validity of the English proficiency test was established through expert judgment, involving two experts in English language teaching and one specialist in language assessment. Each item was reviewed for its alignment with CEFR descriptors and instructional objectives. The resulting Aiken's V coefficients exceeded 0.80, indicating strong expert agreement and satisfactory content validity.[27]

Construct validity was examined using item-total correlation analysis, and items with correlation coefficients below 0.30 were revised or removed. Reliability was assessed using Cronbach's alpha, yielding coefficients above 0.70 for both the pretest and posttest. These values indicate acceptable internal consistency and are consistent with current educational measurement standards.[22], [28]

6. Procedure for the Experiment and Data Collection

The research procedure consisted of three sequential stages: preparation, implementation, and data collection. During the preparation stage, the research instruments were developed, validated, and pilot tested. In the implementation stage, all participants completed the pretest to establish baseline English proficiency. Subsequently, students participated in a Duolingo-based learning intervention over a predetermined instructional period, during which learning activities were conducted autonomously. Throughout the intervention, learners' engagement data were continuously recorded through the Duolingo application.

Upon completion of the intervention, participants completed the posttest under identical

testing conditions to ensure comparability of scores. The data collected for analysis included pretest scores, posttest scores, and individual usage intensity data derived from the usage logs.

7. Data Analysis and ANCOVA Procedure

Data analysis was conducted using Analysis of Covariance (ANCOVA) to examine differences in post-intervention English achievement across usage intensity groups while statistically controlling for initial proficiency. In the ANCOVA model, the pretest score was specified as the covariate, the posttest score served as the dependent variable, and usage intensity level (low, moderate, high) functioned as the fixed grouping factor. This approach allowed for the comparison of adjusted posttest means, thereby reducing bias associated with baseline differences in learners' proficiency.

Prior to conducting ANCOVA, all underlying assumptions were systematically tested. The normality of residuals was examined using Q–Q plots, and the homogeneity of variances was assessed through Levene's test. The linearity between the covariate (pretest) and the dependent variable (posttest) was evaluated using residual scatterplots, while the homogeneity of regression slopes was tested to confirm that the relationship between pretest and posttest scores was consistent across usage intensity groups. All assumptions were satisfied, indicating that the data met the requirements for ANCOVA.

To ensure consistency and transparency in reporting, effect sizes were consistently reported using partial eta squared (η^2) to indicate the magnitude of the observed effects. Statistical significance was evaluated at the conventional alpha level ($p < .05$). This analytical procedure is in line with established recommendations for quasi-experimental pretest–posttest designs in educational research.[26]

RESULT AND DISCUSSION

1. Descriptive Patterns of English Achievement Across Usage Intensity Levels

Table 2. provides an analysis of post-English achievement in the low-, moderate-, and high-usage intensity groups. The descriptive results showed that the Low-intensity group was the most successful and the most consistent in a way ($M = 80.40$; $SD = 5.275$; $CV = 0.066$), the Medium-intensity group was in the middle about the score dispersion ($M = 72.95$; $SD = 7.670$; $CV = 0.105$), and the High-intensity group was the least and the most scattered ($M = 70.00$; $SD = 8.177$; $CV = 0.117$). Thus, the findings suggest that not only are usage levels critical in determining the post-test outcome, but also other conditions, such as learners' initial skills, the quality of their involvement, and the technique of digital tool use.

Table 2. Post-Test Descriptive Analysis by Usage Intensity Level

Usage Intensity Level	N	Mean	SD	SE	Coefficient of variation
High	31	70.00	8.177	1.469	0.117
Moderate	19	72.95	7.670	1.760	0.105
Low	10	80.40	5.275	1.668	0.066

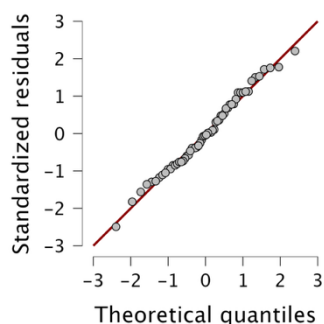
2. Evaluation of Statistical Assumptions for ANCOVA

Table 3 provides the outcome assessments, indicating that the data meet the requirements for homogeneity and normality, allowing for the conduct of an analysis of covariance (ANCOVA). The variance of data among groups was found to be homogeneous, as Levene's test yielded a non-significant value ($F = 1.845$, $p = .167.167$). Hopefully, the Q–Q plot has shown that there was no departure from the regular distribution pattern of residuals, as the standardized residuals were very

close to the theoretical line. Therefore, altogether, these checks have allowed the application of ANCOVA to the study's data to be considered genuine.

Table 3. Test for Equality of Variances (Levene's)

F	df1	df2	P
1.845	2.000	57.00	.167



Pic. 1. Evaluation of Normality Assumption

3. Adjusted Effect of the Duolingo-Based Intervention on English Achievement

Table 4 presents the results of the ANCOVA statistics, which reveal how the Duolingo-powered learning program promotes English attainment while controlling for the initial proficiency of the learners (pre-test scores). The results indicate that the Usage Intensity Level has a statistically significant impact on post-test performance ($F = 4.421$, $p = .016$; $\eta^2 = 0.128$), suggesting that engagement level may be a reason for students' learning gains. The pre-test scores are also a significant, yet smaller factor in the process ($F = 4.258$, $p = .044$; $\eta^2 = 0.062$), thereby emphasizing the continuing role of the starting point proficiency. Furthermore, the high amount of residual variance ($SS = 3081.1$) suggests that other unmeasured factors still play a role in learning outcomes. Taken together, these outcomes provide robust evidence that usage intensity and initial proficiency are among the factors that support students' progress after the intervention.

Table 4. ANCOVA Results for Post-Test Performance

Cases	Sum of Squares	df	Mean Square	F	p	η^2	95% CI	
							Lower	Upper
Usage Intensity Level	486.4	2	243.21	4.421	.016	0.128	0.002	0.287
Pre-test	234.3	1	234.28	4.258	.044	0.062	0.000	0.214
Residuals	3081.1	56	55.02					

4. Pairwise Differences in English Achievement Based on Usage Intensity

Table 5 presents the comparison between pairs of Usage Intensity groups, indicating that no connection exists between them and highlighting how varying engagement levels with Duolingo

can result in different performance outcomes. Post hoc comparisons revealed that only the High- and Low-intensity groups were significantly different in post-test results, with the High group scoring considerably lower than the Low group (Mean Difference = -8.080 , $t(56) = -2.765$, $p_{\text{tukey}} = .021$, $p_{\text{bonf}} = .023$), even after corrections for multiple comparisons. This result indicates that lower usage intensity was associated with higher post-test performance. Contrastingly, none of the High vs. Moderate and Moderate vs. Low comparisons is statistically significant; this is the result that is followed by larger p-values under both the Tukey and the Bonferroni corrections. These results imply that students with high and low usage levels exhibit the most incredible Difference in results, with high users being more likely to achieve a high post-test score.

Table 5. Pairwise Group Differences Based on Usage Intensity.

		Mean Difference	SE	df	T	p_{tukey}	p_{bonf}
High	Moderate	-3.652	2.188	56	-1.669	.226	.302
	Low	-8.080	2.922	56	-2.765	.021	.023
Moderate	Low	-4.429	3.247	56	-1.364	.367	.534

5. Comparison of Adjusted Post-Test Means Across Usage Intensity Groups

Table 6 presents a comparison of adjusted posttest means after controlling for initial English proficiency. The results clarify significant differences in English achievement associated with variations in students' usage intensity and engagement with the application. Adjusted means are categorized based on the three levels of technology use; very low adjusted mean of ($M = 70.16$, 95% CI [67.49, 72.84]) was the High-intensity group, the second highest adjusted mean was recorded in the Moderate-intensity group ($M = 73.82$, 95% CI [70.30, 77.33]), and the highest adjusted mean was in the Low-intensity group with the widest confidence interval ($M = 78.24$, 95% CI [73.10, 83.39]). This area is so broad that it encompasses almost all the higher and lower means of the other groups. The significance of the t-values, which are strongly negatively correlated within each group ($p < 0.001$), indicates that the means are very significantly different among all groups. Consequently, the pattern of the results collectively suggests a relationship between lower technology usage and higher adjusted post-test performance, while also emphasizing moderated engagement as a contributor to the predictive factor.

Table 6. Comparison of Adjusted Post-Test Means Across Usage Intensity Groups

Usage Intensity Level	Marginal Mean	95% CI for Mean Difference		SE	T	df	P
		Lower	Upper				
High	70.16	67.49	72.84	1.335	52.57	56	< .001
Moderate	73.82	70.30	77.33	1.753	42.11	56	< .001
Low	78.24	73.10	83.39	2.568	30.47	56	< .001

Discussions

This study aimed to investigate the utility of a Duolingo-based learning intervention in enhancing students' English achievement and the role of intensity in usage as a predictor, while controlling for students' original levels of proficiency. The ANCOVA results uncovered two main findings. To begin with, when controlling for pre-test scores, the Duolingo-based learning intervention was very effective, as it significantly raised the students' English achievement. This supports the theory that a combination of traditional learning and mobile-assisted learning is a step forward in language development. Secondly, the intensity of usage—as quantified by measures of session frequency, duration, and cumulative time—proved to be a reliable predictor of the students'

performance after the intervention. However, the relationship was not straightforward: students who used "moderately to low" usage intensity achieved higher adjusted post-test scores compared to those with high usage intensity. Importantly, although the low-intensity group obtained the highest raw posttest mean, the primary interpretation is based on ANCOVA-adjusted posttest means, which revealed a statistically significant effect of usage intensity ($F(2,57) = 4.421, p = .016, \eta^2 = .128$). This clarifies that higher usage intensity does not necessarily translate into superior learning gains once initial proficiency is controlled.

In response to the first research question, the results clearly reveal that the intervention using Duolingo has significantly increased the English level of the learners, even after adjusting for the pre-test scores. This makes it possible that the rise in test results is not just due to the prior knowledge students have, but also to the regular practice of Duolingo's materials. One of the reasons behind the improvement has been the gamification, repetition, and adaptive feedback. [3], [15], [16], [30] The evidence also shows that the learning process through the app can be beneficial in a way that it is a real match for the traditional classroom instruction, which further supports the previous type of analysis that pointed to the good consequences of the MALL regarding English proficiency. [1], [5]

The way participants engage in the virtual sphere in mobile learning has been considered a significant variable in achieving the desired learning outcome. Paradoxically, it is very intense usage that is associated with low achievement levels—if not reverse; this above-average involvement may invite nothing else but mental exhaustion and superficial processing. This interpretation is directly supported by the pairwise comparison results, which showed that the high-intensity group performed significantly lower than the low-intensity group (Mean Difference = $-8.080, p < .05$), providing empirical evidence of a non-linear relationship between usage intensity and learning outcomes.

The learners who are committed to their smartphones all the time and are using an extensive number of certain apps might do that for one or more of the following reasons: they are chasing after the points to compete with others, or they want to see that they have an unbroken chain, which is really extrinsic motivation. In this way, they are missing out on the real gain, which comes from doing something more profound and meaningful. On the other hand, users who use their smartphones moderately or sparingly are the ones most likely to have a more profound and practical engagement with the content, as they do not become too tired or bored with it and can thus engage in reflective learning. Post-test results for these two groups were significantly better. The observed effect size ($\eta^2 = .128$) indicates that usage intensity accounts for a meaningful proportion of variance in post-intervention English achievement, reinforcing the interpretation that excessive engagement yields diminishing returns rather than additive benefits. Although the "optimal engagement" hypothesis from gamified learning theory suggests that instruction should shift the balance between motivation and metacognition, it actually outlines non-extreme points. In the present study, this optimal engagement pattern is statistically reflected in the superior adjusted outcomes of the moderate-to-low usage groups, rather than the high-intensity group.

The results of this study are in concordance with the standard proof that the process of learning via Duolingo can amplify different parts of English language skills, among them vocabulary acquisition [18], reading comprehension [31], and speaking performance. [6], [24] In the same way that was done in previous investigations, the current research has proven that the use of playful learning with a game-like element can lead to noticeable improvements in skills. [8], [13]

However, this research presents a novel perspective by demonstrating that excessive app usage can actually be detrimental to one's performance, providing a counterpoint to the widespread belief that increased usage leads to greater success. Such a non-linear relationship sheds light on digital study habits, and the quality of engagement appears to be more important than its quantity,

as [29] and the mediating effects of motivation and self-regulation in MALL contexts, described by [12], are supported by the findings. The diminished adjusted performance of high-intensity users supports the notion of diminishing returns, consistent with theoretical assumptions of cognitive load and self-regulation in technology-enhanced learning.

Moreover, the research also shows that the use of analytic methods can be a valid method for educational technology research. Similarly, the [3] study has found that there was a positive correlation between the use of apps and learners who showed measurable gains in their L2 skills; the newly completed research, nonetheless, argues that intensity of use could be a behavioral predictor of the student's success in digital environments. However, by demonstrating the performance decline among super-users of the technology, this study broadens current thinking by incorporating the law of diminishing returns into mobile language learning.

This research study has advanced the conceptual understanding of MALL, establishing, in principle, and through the collection of appropriate data, that "usage intensity" is a measurable variable predicting language performance. The study, according to [10] and [3], suggests that models of exposure should be abandoned in favor of frameworks that recognize behavior, cognition, motivation, self-regulation, and engagement quality as mediation variables. Additionally, this study's success in evaluating the curvilinear pattern of engagement lends support to cognitive load theory and self-determination theory, which suggest that the most effective learning outcomes are achieved when individuals reach the point of optimal engagement, rather than engaging to the maximum level.

Educators, policymakers, app developers, and other stakeholders benefit from these research findings. It is observed that in the case of educators, Duolingo should be utilized as a well-structured and supplementary tool, with the support of clearly defined time frames (e.g., 20-30 minutes a day), to ensure that engagement does not exceed the optimal cognitive threshold. Policymakers can utilize the insight here to adapt a curriculum based on data that takes into account mobile learning, while also avoiding an unhealthy addiction to game-based rewards. The developers perceive this research as a valuable contribution to the broader discussion of technology-enhanced language learning and user engagement.[32] However, the present study does not directly examine or evaluate artificial intelligence mechanisms, focusing instead on learning outcomes and usage patterns.

Moreover, the incorporation of usage analytics dashboards provides a way for educators to identify learners who are at risk of losing interest or becoming too engaged. When detected, these are signs that the instructors should take pedagogical action at the right time.[2], [5] The study proposed a balanced engagement approach, which can be seen as a contributor to improvements in app-based instruction in vast EFL contexts.

CONCLUSION AND IMPLICATIONS

The findings demonstrate a statistically significant effect of the Duolingo-based intervention on students' English achievement after controlling for initial proficiency. The adjusted analysis indicated meaningful differences across Usage Intensity groups, with post hoc comparisons revealing a significant difference between high- and low-intensity users (Mean Difference = -8.080 , $t(56) = -2.765$, $p < .05$). These results suggest that Usage Intensity functions as a significant yet non-linear predictor of learning outcomes. Notably, moderate engagement levels were associated with comparatively higher performance, whereas excessive usage did not correspond to improved achievement. Collectively, the findings challenge the assumption that increased usage necessarily leads to better outcomes and instead highlight the importance of balanced engagement in technology-assisted language learning.

The study, from a theoretical angle, has highlighted the subject of MALL, focusing on both user performance and interactional data, thereby enabling us to understand which framings would be most suitable for the next generation. Practically, it has provided a roadmap for teachers, IQ test administrators, and community centers to facilitate easy learning of modern IT, the cloud, and staying online.

Data suggest that, despite digital platforms taking over our time, the way of language learning to come is not only through technological means but also through human-centered, reflective, and intentional methods. Next, linking engagement, cognition, and attainment will pave the way for a research-backed, AI-empowered language education system where the brightest and most proficient students will be at its core.

BIBLIOGRAPHY

- [1] M. Mihaylova, S. Gorin, T. P. Reber, and N. Rothen, "A Meta-Analysis on Mobile-Assisted Language Learning Applications: Benefits and Risks," *Psychol Belg*, vol. 62, no. 1, pp. 252, 2022, <https://doi.org/10.5334/pb.1146>.
- [2] K. Karakaya and A. Bozkurt, "Mobile-assisted language learning (MALL) research trends and patterns through bibliometric analysis: Empowering language learners through ubiquitous educational technologies," *System*, vol. 110, pp. 102925, 2022, <https://doi.org/10.1016/j.system.2022.102925>.
- [3] M. Kessler, S. Loewen, and T. Gönülal, "Mobile-assisted language learning with Babbel and Duolingo: comparing L2 learning gains and user experience," *Comput Assist Lang Learn*, vol. 38, no. 4, pp. 690–714, 2025, <https://doi.org/10.1080/09588221.2023.2215294>.
- [4] B. T. Suryanto, "The Impact of Using DUOLINGO Application on Students' English Learning Motivation," *International Journal of English Education and Linguistics (IJoEEL)*, vol. 6, no. 2, pp. 200–211, 2024, <https://doi.org/10.33650/ijoeel.v6i2.9700>.
- [5] Mujiono, "Mobile-Assisted Language Learning Intervention and Its Effect on English Language Proficiency of EFL Learners: A Meta-Analysis," *Theory and Practice in Language Studies*, vol. 13, no. 5, pp. 1124–1135, 2023, <https://doi.org/10.17507/tpls.1305.05>.
- [6] I. Y. Kazu and M. Kuvvetli, "Improve speaking skills with Duolingo's mobile game-based language learning," *Asian Journal of Education and Training*, vol. 10, no. 1, pp. 62–75, 2024, <https://doi.org/10.20448/edu.v10i1.5488>.
- [7] A. Alqarni, "Effect of Mobile Assisted Learning on English Language Vocabulary and Grammar: The Saudi Arabian Context as a Case Study," *Arab World English Journal*, no. 10, pp. 246–265, 2024, <https://doi.org/10.24093/awej/call10.16>.
- [8] M. Shortt, S. Tilak, I. Kuznetcova, B. Martens, and B. Akinkuolie, "Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020," *Comput Assist Lang Learn*, vol. 36, no. 3, pp. 517–554, 2023, <https://doi.org/10.1080/09588221.2021.1933540>.
- [9] V. Handayani and I. Damayanti, "Mobile-Assisted Language Learning for Vocabulary Development: Insights from Indonesian Pre-Service English Teachers Using HelloTalk," *Jurnal Taburi*, vol. 19, no. 2, pp. 99–114, 2022, <https://doi.org/10.30598/tahurivol19issue2page99-114>.
- [10] S. Liu, S. Zhang, and Y. Dai, "Do mobile games improve language learning? A meta-analysis," *Comput Assist Lang Learn*, pp. 1–29, 2025, <https://doi.org/10.1080/09588221.2025.2528786>.

- [11] Mujiono and S. Fatimah, "Moodle Integration Intervention in EFL Virtual Classroom and Academic Flow on University Students' Achievement in Writing," *Theory and Practice in Language Studies*, vol. 12, no. 10, pp. 2182–2190, 2022, doi: 10.17507/tpls.1210.26.
- [12] A. Y. Purwaningrum, N. S. Lengkanawati, and F. N. Yusuf, "Examining Mobile-assisted Language Learning as an Autonomous Writing Tool for Indonesian Secondary EFL Students," *Register Journal*, vol. 17, no. 2, pp. 328–351, 2024, <https://doi.org/10.18326/register.v17i2.328-351>.
- [13] F. Fitriani and G. P. Mokodompit, "The Effect of Duolingo Application in Improving Students' English Skills: A Research of Grade X Students at SMKN 3 Baubau," *JELITA*, vol. 6, no. 1, pp. 50–60, 2024, <https://doi.org/10.56185/jelita.v6i1.886>.
- [14] V. Liebni, Indra Perdana, and Elanneri Karani, "Expeditencies of Duolingo in English Language Learning: A Systematic Literature Review," *English Journal of IndragirI*, vol. 9, no. 2, pp. 453–462, 2025, <https://doi.org/10.61672/eji.v9i2.2997>.
- [15] B. Smith, X. Jiang, and R. Peters, "The effectiveness of Duolingo in developing receptive and productive language knowledge and proficiency," *Language Learning & Technology*, vol. 28, no. 1, pp. 1–26, 2024, <https://doi.org/10.64152/10125/73595>.
- [16] E. Sudina and L. Plonsky, "The effects of frequency, duration, and intensity on L2 learning through Duolingo," *Journal of Second Language Studies*, vol. 7, no. 1, pp. 1–43, 2024, <https://doi.org/10.1075/jsls.00021.plo>.
- [17] R Vesselinov, "The Babbel efficacy study. Final report, 2022".
- [18] M. Febrianti, I. N. Rahmawati, A. Aisyah, A. K. P. Nasution, and A. Aprianto, "Using the Duolingo Application as a Vocabulary Learning Tool in Higher Education," *Journal International of Lingua and Technology*, vol. 3, no. 2, pp. 345–361, 2024, <https://doi.org/10.55849/jiltech.v3i2.673>.
- [19] R. H. Mogavi, B. Guo, Y. Zhang, E.-U. Haq, P. Hui, and X. Ma, "When Gamification Spoils Your Learning: A Qualitative Case Study of Gamification Misuse in a Language-Learning App," Mar. 2022, <https://doi.org/10.1145/3491140.3528274>.
- [20] T. R. Knapp, "Why Is the One-Group Pretest–Posttest Design Still Used?," *Clin Nurs Res*, vol. 25, no. 5, pp. 467–472, 2016, <https://doi.org/10.1177/1054773816666280>.
- [21] C. W. Johnson, "A More Rigorous QUASI-Experimental Alternative to the One-Group Pretest-Posttest Design," *Educ Psychol Meas*, vol. 46, no. 3, pp. 585–591, 1986, doi: 10.1177/0013164486463011.
- [22] N. Dehghan Nayeri, F. A. Noodeh, H. S. Nia, A. Yaghoobzadeh, K. A. Allen, and A. H. Goudarzian, "Statistical Procedures Used in Pretest-Posttest Control Group Design: A Review of Papers in Five Iranian Journals," *Acta Med Iran*, 2024, <https://doi.org/10.18502/acta.v61i10.15657>.
- [23] J. R. Rachels and A. J. Rockinson-Szapkiw, "The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy," *Comput Assist Lang Learn*, vol. 31, no. 1–2, pp. 72–89, 2018, <https://doi.org/10.1080/09588221.2017.1382536>.
- [24] D. S. Amaniarsih, N. Asrul, J. Juliana, and E. Ginting, "The Effectiveness Of Using The Duolingo Application In Improving Students' Motivation And English-Speaking Skill At Universitas Potensi Utama Medan," *Pedagogi: Jurnal Ilmiah Pendidikan*, vol. 11, no. 1, pp. 46–54, 2025, <https://doi.org/10.47662/pedagogi.v11i1.933>.

- [25] D. Asfarini, B. Basuki, A. Thien Wan, and E. S. Masykuri, "Review Study Learning Apps in Teaching Vocabulary Using Duolingo," *Scripta: English Department Journal*, vol. 11, no. 1, pp. 48–56, 2024, <https://doi.org/10.37729/scripta.v11i1.3836>.
- [26] E. Hong, "Quasi-Experimentation: Two Group Design," in *International Encyclopedia of Education*, Elsevier, 2010, pp. 128–133. <https://doi.org/10.1016/B978-0-08-044894-7.01686-9>.
- [27] I. N. Aziz and Y. A. S. Dewi, "The Implementation of Contextual Teaching and Learning on English Grammar Competence," *Alsuna: Journal of Arabic and English Language*, vol. 2, no. 2, pp. 67–95, 2019, <https://doi.org/10.31538/alsuna.v2i2.392>.
- [28] M. J. Uddin, B. N. Panda, and P. C. Agarwal, "Detective Learning of the Concepts of Acids, Bases and Salts in Physical Science by Ninth-grade Students using a Metacognitive Instructional Strategy: A Quasi-Experimental Study," *Journal of Education, Society and Behavioural Science*, pp. 96–107, 2022, <https://doi.org/10.9734/jesbs/2022/v35i121199>.
- [29] A. Lu, S. Liu, and W. Chen, "Perceived usefulness of English (L2) learning apps and language mindset mediated by flow and motivation intensity: a serial mediation model and a network analysis," *Interactive Learning Environments*, pp. 1–20, 2024, <https://doi.org/10.1080/10494820.2024.2375637>.
- [30] X. Jiang, R. Peters, L. Plonsky, and B. Pajak, "The Effectiveness of Duolingo English Courses in Developing Reading and Listening Proficiency," *CALICO Journal*, vol. 41, no. 3, pp. 249–272, 2024, <https://doi.org/10.1558/cj.26704>.
- [31] P. Ajisoko, "Using Duolingo apps to improve English reading comprehension of engineering students in Universitas Borneo Tarakan". *Exposure: Jurnal Pendidikan Bahasa Inggris*. Vol. 11, No. 1, pp. 1-6, 2022, <https://doi.org/10.26618/exposure.v11i1.6452>
- [32] J. Burstein and G. T. LaFlair, "Where Assessment Validation and Responsible AI Meet," Vol. 50, Special Issue, pp. 120-137, 2025. <https://doi.org/10.32038/ltrq.2025.50.09>