

Enhancing Lexical Development through Innovative Vocabulary Instruction in a Secondary EFL Classroom: A Quasi-Experimental Study

Mejora del desarrollo léxico mediante una instrucción innovadora de vocabulario en un aula de inglés como lengua extranjera de secundaria: un estudio cuasi-experimental

Narcisa de Jesús Gómez León, Juanita Argudo-Serrano

Abstract

This quasi-experimental study examines the impact of some vocabulary strategies on the lexical development of first-year high school students learning English as a foreign language (EFL) at an educational institution in the Ecuadorian Amazon. Although language learning supports reading comprehension, communication, and general linguistic competence (Nation, 2013; Read, 2000; Schmitt, 2010; Webb & Nation, 2017), it is often taught as a memorization of isolated words. This study involved a three-week instructional intervention that included retrieval practice, spaced repetition, and vocabulary activities designed to promote students' lexical processing. The study employed a quantitative quasi-experimental design with 48 first-year high school students divided into two groups: an intervention group and a control group. Both groups were administered the Updated Vocabulary Levels Test, developed by Webb et al. (2017), with Form A used as a pretest and Form B as a posttest. The results showed that the experimental group scored significantly higher on the posttest than the control group at all vocabulary levels and in the total score. These findings suggest that implementing retrieval, repetition, and contextualized activities can promote the development of receptive vocabulary in secondary school English as a foreign language class, although this study is not considered conclusive causal evidence.

Keywords: Vocabulary learning, EFL, retrieval practice, secondary education, lexical development.

Narcisa de Jesús Gómez León

Universidad Católica de Cuenca | Cuenca | Ecuador | narcisa.gomez21@est.ucacue.edu.ec
<https://orcid.org/0009-0004-2107-7901>

Juanita Argudo-Serrano

Universidad Católica de Cuenca | Cuenca | Ecuador | juanita.argudo@ucacue.edu.ec
<http://orcid.org/0000-0002-3337-7803>

Resumen

En este estudio cuasiexperimental se examinó el impacto de algunas estrategias innovadoras de vocabulario en el desarrollo léxico de estudiantes que cursan el primer año de Bachillerato que aprendían inglés como lengua extranjera (EFL) en una institución educativa de la Amazonía ecuatoriana. A pesar de que el aprendizaje de idiomas apoya la comprensión lectora, la comunicación y la competencia lingüística general (Nation, 2013; Read, 2000; Schmitt, 2010; Webb y Nation, 2017), a menudo se lo enseña como una lista de memorización de palabras aisladas. En este trabajo se realizó una intervención instruccional de tres semanas que incluía la práctica de recuperación, la repetición espaciada y las actividades de vocabulario creadas para promover el procesamiento léxico del estudiante. El estudio trabajó con un diseño cuasiexperimental cuantitativo con 48 estudiantes de primer año de Bachillerato dividido en dos grupos, uno de intervención y otro de control. A ambos grupos se aplicó la Prueba de Niveles de Vocabulario Actualizada, desarrollada por Webb et al. (2017), administrándose la Forma A como pretest y la Forma B como postest. Los resultados obtenidos mostraron que el grupo experimental obtuvo puntuaciones significativamente más altas en el postest que el grupo de control en todos los niveles de vocabulario y en la puntuación total. Estos hallazgos muestran que poner en práctica la recuperación, la repetición y las actividades contextualizadas pueden favorecer El desarrollo del vocabulario receptivo en las clases de inglés como lengua extranjera en la educación secundaria, aunque este trabajo no se considere como una prueba causal concluyente.

Palabras clave: Aprendizaje de vocabulario, educación secundaria, desarrollo léxico, inglés como lengua extranjera.

Introduction

Vocabulary is an important element in learning English as a foreign language (EFL) and teaching it in secondary education is no exception. It is well known that vocabulary helps learners understand what they read, express their ideas, and confidently begin using the second language in the classroom (Nation, 2013; Read, 2000; Schmitt, 2010; Webb & Nation, 2017). However, in recent years it has been observed that teaching vocabulary in isolation, using purely rote memorization methods, is detrimental rather than beneficial. Therefore, it is important to pay special attention to how students are repeatedly exposed to new words and how they use them (Nation, 2013; Webb & Nation, 2017). However, this doesn't mean that all memorization is wrong. It's important to highlight the subtle yet effective elements of flexible use so that what has been retained in practice functions in long-term memory.

On the other hand, if exposure outside the EFL classroom is limited, the acquisition of new vocabulary has few opportunities, and the entire burden of learning falls on the school. Therefore, classrooms are where students have the opportunity to learn and internalize what they learn. In this regard, studies such as those by Nation (2013), Schmitt (2010), and Webb and Nation (2017), have shown that learning new words requires repetitive, yet meaningful, encounters. In other words, learning must be contextualized to develop a deep understanding and apply those words communicatively, attempting to use them freely.

One of the challenges teachers face is to know how to help students recognize and understand vocabulary components in applied contexts. While students can certainly recognize new words in classroom activities at a given moment, problems often arise when they try to use them

in situations outside the classroom. Interaction with vocabulary questions creates a particular problem. Some research on vocabulary assessment raises concerns for teachers because time is limited, yet curricular expectations are high (Nation, 2001; Read, 2000; Schmitt, 2010). This situation demonstrates that simply repeating word lists does not foster communicative processes that extend beyond the classroom. Therefore, in order to obtain success, communicative tasks that simulate real-life situations are necessary, but this often leads to a lack of confidence among students. When students understand a word, they not only face challenges but are also able to tackle more complex texts (Coxhead, 2000). In this sense, designing flexible activities to interpret meaning and express the student's ideas may be the most appropriate path (Coxhead, 2000; Nation, 2013; Schmitt, 2010).

Another problem teachers often report on is the limitation of available time and resources. Those who have studied this problem have generated significant evidence on repetition, remediation, and exposure, but only a small number of studies have actually produced reports on short-term interventions in classrooms where English is taught as a foreign language, particularly in Ecuador. We have reviewed literature on remediation practices, identifying quasi-experimental studies in master's and undergraduate theses, in particular. However, none of them focused on secondary school students, so this study could contribute to the research of this age group of students, who are around 15 years old.

As Hacking and Tschirner (2017), point out, meaningful language skills can build confidence in students who struggle to communicate due to shyness or embarrassment about being observed, a common issue during adolescence. Therefore, attention should be paid to these skills, in which vocabulary plays a crucial role. Developing a deep lexical knowledge fosters the ability to handle complex situations, such as real-life communication. This research calls for innovative approaches that consider aspects of personality psychology and adolescence.

Indeed, the psychology of second language learning posits that pedagogical processes should be active and employ an appropriate pace of introducing the new language, and therefore, new vocabulary. Studies conducted abroad, such as those by Karpicke and Roediger (2008), Laufer and Hulstijn (2001), Roediger and Karpicke (2006), and especially those by Webb (2007) and Webb and Nation (2017), have tested hypotheses regarding strategies such as retrieval, spaced repetition, and the use of new vocabulary in more contextualized environments. These studies have demonstrated, to varying degrees, that timely memory retention occurs by activating the student's memory. Therefore, it is important to determine whether specific exercises can be implemented in a classroom within the Latin American context, particularly in the Ecuadorian Amazon where this study is being conducted.

Therefore, the problem that concerns this work is the lack of contextualized evidence on new vocabulary instruction to demonstrate whether meaningful and repeated opportunities can generate appropriate knowledge, at least for a real classroom setting. The research of Karpicke and Roediger (2008), Roediger and Karpicke (2006), Webb (2007), and Webb and Nation (2017)

serve as inspiration for this work, as it provides access to literature on the functioning of specific strategies applied in other contexts.

Our purpose is to examine the impact of a vocabulary instruction intervention based on retrieval practice, repetition, and contextualized activities to develop new vocabulary in first-year high school students in a secondary school in Ecuador. In this regard, we pose at least two research questions. The first asks: What differences were observed in receptive vocabulary scores between the experimental and control groups at the pre-test and post-test times? And the second question asks: What differences were observed between the experimental and control groups in vocabulary frequency levels as measured by the Updated Vocabulary Levels Test?

Theoretical Framework and Literature Review

Theoretical Framework

The acquisition of new vocabulary within a new language, as Nation (2001), Read (2000), and Schmitt (2010) point out, is a complex process that requires focusing on several dimensions. In this regard, the most recent literature indicates that while it is true that forms must be considered, it is equally important to understand their meaning, that is, their grammatical behavior, taking into account associations, frequency, register, and correct use within the context. It is important to note that there are efficient vocabulary teaching methods that make the most of the limited time allotted to foreign language instruction. This allows for the promotion of both receptive comprehension and productive use once the student is outside the classroom.

Now, on a psychological level, cognition must be situated for the learning process in order to be productive (Ellis, 2002). Laufer and Hulstijn (2001), argue that tasks require activating lexical elements that generate better retention than traditional rote learning; that is, students must learn to make decisions in order to grasp the meaning, form, and use of new words. Thus, they suggest activities such as sentence construction, contextual use, and guided production, which allow for greater long-term improvement. These findings point to a pre-pedagogical approach that promotes more spontaneous student participation.

Learning new vocabulary requires spontaneous and ongoing participation of students. In this regard, Craik and Lockhart (1972) assert that information that has been processed is more likely to be retained in long-term memory. It should be noted that context provides the opportunity for deep processing to analyze meaning, establish associations, and use the new vocabulary in other scenarios. An example of this is the approach developed by Joe (1995), who, to generate sentences with new words, based on personal experiences, which led to greater promotion of lexical items and increased cognitive development. Consequently, teaching deep processing involves identifying and applying effective practices in which the individual is compelled to use a word.

Retrieving a word from memory is of particular interest in this work, as it is an effective mechanism for consolidating new vocabulary. In this regard, the research conducted by Karpicke

and Roediger (2008), Roediger and Karpicke (2006) show that remembering is also an exercise that must be strengthened by introducing systematic opportunities in class to retrieve new vocabulary, for which the teacher must schedule when to use those words that were taught in the past again so that the student has the opportunity to search for and retrieve them easily in his memory.

Learning new vocabulary requires repetition exercises that strengthen lexical knowledge. While this activity seems to be demonized by contemporary educators, reiterative exposure to words helps students consolidate their understanding of meaning, forms, and uses. However, there is no other way to ensure assimilation than through isolated repeated use. Now, it's not about mindlessly repeating words. According to Nation (2001, 2013), when words are encountered in different contexts, students end up acquiring a deeper and more flexible vocabulary. In other words, new vocabulary must be used in many situations so that the meaning is assimilated without needing a translated equivalent. Thus, repetition becomes meaningful and effective; in other words, it's an interaction with the target language. Activities such as writing sentences, participating in debates, or interacting with peers or the teacher, Webb (2005) points out, can generate retention and facilitate the transition from receptive to active knowledge of vocabulary. Nation (2013) states that repeating a new word is an ineffective task unless students encounter that word in various contexts. Karpicke and Roediger (2008) argue that retrieval strengthens long-term memory because it involves active recall instead of the mindless repetition common in the past. Therefore, repetition must be productive and combined with complementary, situated strategies.

Literature Review (State of the Art)

The most recent studies show positive effects of vocabulary retrieval and retention practice (Li et al., 2024; Qu et al., 2024). Those who study the effectiveness of recall in strengthening long-term retention and vocabulary retrieval show that classroom exercises are effective (Karpicke & Roediger, 2008; Roediger & Karpicke, 2006). A recurring problem with this evidence is that it is conducted under controlled experimental conditions, which certainly do not encompass all the variables involved in an educational environment far removed from the location where the experiment was applied. However, experimentation is often the best way to find evidence. Therefore, it is necessary to control complementary elements such as student motivation, sufficient time and scheduling, and a classroom environment where instructional and practice strategies flow and where students have developed a sufficiently consolidated relationship to facilitate interaction (Ushioda, 2021).

Repeating vocabulary, as we have mentioned, shows that significant improvement in lexical retention is possible. As we have already stated, the most prominent evidence is possibly that of Webb (2007), who demonstrated that repeated exposure to lexical items can significantly increase retention, leading to a deeper understanding of the meaning of new words. It is also important to note that retrieval learning improves long-term memory, according to Karpicke and Roediger (2008) and Roediger and Karpicke (2006). The findings of these authors show that active recall

is a relevant mechanism in vocabulary acquisition, highlighting the idea that effective vocabulary learning requires active participation rather than passive exposure.

Instruction plays an important role, and some studies have analyzed the effectiveness of EFL exposure in situated contexts. The study by Webb and Chang (2015) revealed that second language vocabulary learning through intensive reading and audio support generated evidence of the impact of the frequency and distribution of new words on learning. Their results demonstrated that repeated exposure and the way in which new words are accessed lead to much more effective learning. Webb and Nation (2017) have shown that repetition is not inherently bad; in fact, repeated encounters, attention to meaning, and the way in which new vocabulary is introduced create conditions for meaningful use that can be extrapolated to other contexts beyond the classroom. Furthermore, Webb (2007) has demonstrated that vocabulary development is related to specific learning periods; that is, it is possible to implement short didactic interventions, provided they are carefully structured in terms of how they are accessed and how frequently they are carried out.

While research on vocabulary has certainly increased, there are few quasi-experimental studies conducted in Ecuador. For example, controlled studies using this method generally come from foreign researchers who have taken advantage of the most advanced academic and technological resources, as is the case with Godwin-Jones (2020) and Webb and Chang (2015). In fact, they have reflected their own reality within the practice of student groups. Therefore, this evidence is far removed from the Ecuadorian context, but it offers important background information for our own research. For this reason, we need to examine how to implement innovative strategies in Ecuadorian classrooms.

Webb (2007) suggests that combining retrieval and repetition to acquire new vocabulary, with repeated encounters with the lexical item, allows knowledge to be progressively integrated and move toward use. Nation (2013) points out that repetition trains memory and makes it easier to retain information long-term, but this requires working with meaningful experiences rather than rote learning.

Authors such as Nation (2001), Read (2000), and Schmitt (2010) have demonstrated that receptive and productive vocabulary are two different things when it comes to developing new vocabulary. Receiving information—learning to recognize and understand lexical items when reading or listening—is one thing, but producing vocabulary—speaking or writing using those new words—is another. In this sense, Laufer and Goldstein (2004) assert that they are not only two different things, but that one goes beyond the other. The productive aspect of vocabulary requires learning to recognize vocabulary in order to move beyond receptive learning. Once receptive vocabulary is consolidated, productive vocabulary must be progressively supported. Learning in context involves developing vocabulary by allowing students to interpret its meaning, understand its appropriate use, and begin to connect the new vocabulary with meaningful communicative encounters. For this reason, integrating contextualized activities contributes to the productive aspect of vocabulary development.

Generally, literature highlights the importance of combining retrieval, repetition, and contextual learning, which promotes the acquisition of a deeper vocabulary. Many scholars concur with this view (Karpicke & Roediger, 2008; Nation, 2013; Roediger & Karpicke, 2006; Webb, 2007; Webb & Nation, 2017). While their experiences certainly support our proposal to implement an intervention process in an Amazonian setting, it is necessary to acknowledge what might vary in a moderately controlled situation, such as a quasi-experimental study. Therefore, this study aims to address this gap by investigating the impact of a short-term, classroom-based vocabulary intervention in a secondary EFL context in Ecuador.

Several studies related to the present investigation used some teaching materials to improve vocabulary. For example, Nakata (2011) used flashcards to teach vocabulary in a second language; his work consisted of work with repeated presentations, which help retention of new words. Another study developed by Webb (2007) showed improvement in vocabulary in 121 students during ten classes. His evidence supports the idea of daily routines to consolidate new vocabulary. Karpicke and Roediger (2008) worked with paired words, recovering one from the other word. Their results showed that students who continued to actively recall the words they had learned remembered significantly more vocabulary one week later than those who simply reviewed or studied the flashcards. Komachali and Khodareza (2012) compared students who learned vocabulary using flashcards with a group that received traditional instruction. The group that used flashcards scored significantly higher on the subsequent vocabulary test.

Methods

This study adopts a quantitative quasi-experimental research design (Cabrera-Tenecela, 2023) to examine the impact of a vocabulary instruction intervention on EFL students' lexical development. Therefore, a non-equivalent pretest-posttest control-group design was used to measure changes in students' receptive vocabulary knowledge and to compare the performance of an experimental group and a control group. This approach allows for the systematic comparison of students' performance and the evaluation of the effectiveness of the instructional strategies implemented during this study (Creswell & Creswell, 2018).

The participants were 48 first-year Baccalaureate students, approximately 15 years old, enrolled in EFL classroom groups at a public secondary school in Ecuador. The sample included 25 students from the Technical Baccalaureate program and 23 students from the Science Baccalaureate program. Gender distribution was not collected as part of the dataset. The Technical Baccalaureate group was selected as the experimental group because students in this class showed greater willingness to participate collaboratively in the intervention. The Science Baccalaureate group served as the control group. Both groups were taught by the same teacher. According to the institutional curriculum, students at this level are expected to be working toward the A2/B1 range; however, no independent proficiency test was administered before the intervention. Both groups received regular English instruction as part of the institutional timetable, with 3 English

class periods per week, each lasting approximately 45 minutes. The sample was selected through convenience sampling (Kline, 2017; López, 2004), as it involves an intact group of students within an existing classroom setting. Because the study was conducted in an existing school setting, the participants were selected from intact classes rather than being randomly assigned to newly formed groups. Therefore, convenience sampling was used, which is common in quasi-experimental educational research (Creswell & Creswell, 2018).

Ethical considerations were addressed by obtaining institutional authorization and informing students and their legal representatives about the academic purpose of the study. Participants' confidentiality was protected, and the results were used only for research purposes. All students were informed verbally about the study, and written informed consent was obtained from their parents or legal guardians. The instrument used in this study was the Updated Vocabulary Levels Test developed by Webb et al. (2017). Form A was administered as the pretest and Form B as the posttest. These forms are appropriate because the instrument was developed and validated as two equivalent forms of the Vocabulary Levels Test, which allows researchers to assess students' vocabulary knowledge before and after an instructional intervention without simply repeating the same test. The test measures receptive vocabulary knowledge across word-frequency levels and provides a suitable measure for identifying changes in students' lexical knowledge.

Because the Updated Vocabulary Levels Test primarily measures receptive vocabulary knowledge, the present study focused its statistical analysis on receptive vocabulary development. Contextualized classroom activities were included in the intervention to encourage meaningful use of vocabulary. However, productive vocabulary use was not treated as the main measured outcome because no additional productive vocabulary task or scoring rubric was incorporated into the measurement design.

To ensure the validity and reliability of the instrument, this study used the validated Updated Vocabulary Levels Test rather than a teacher-made test. The selection of Form A and Form B helped maintain consistency between the pretest and posttest while reducing the possibility that students improved only because they have already seen the same items. The administration conditions were kept as similar as possible in both testing moments.

The study was conducted in three phases: pre-test, intervention, and post-test. Data analysis followed. The pre-test was administered on May 13th to the experimental group and on May 15th to the control group. The Updated Vocabulary Levels Test (Form A) was used for assessment. This assessment took place in the classroom, the students' regular learning environment, allowing for the measurement of their initial receptive vocabulary knowledge.

The intervention was implemented over a three-week period during the students' regular school schedule, from May 14th to June 3rd. Each week, the students in the experimental group were introduced to a new set of words and worked through a sequence of activities designed to be meaningful enough to facilitate vocabulary acquisition. The process began with an initial presentation, accompanied by guided practice. Throughout the class, remedial exercises were applied,

but review was also spaced out, as Webb suggests. The activities were as contextualized as possible, related to sentence construction about what the students thought or felt. This was achieved through writing exercises and oral interaction. Regular review was monitored by a timer, allowing the teacher to keep track of when to reuse each word. While a word might be acquired in one class, it continued to be used systematically in subsequent classes. During the three weeks, the control group continued performing the regular EFL learning activities mandated by the national curriculum, without receiving the structured intervention of new vocabulary.

During the posttest phase, students in both groups completed Form B of the Updated Vocabulary Levels Test to measure changes in receptive vocabulary knowledge following the intervention. This posttest evaluation was done on June 4 and June 5. Finally, during the data analysis phase, pretest and posttest results from the experimental and control groups were compared to evaluate the effectiveness of the instructional intervention. Because the available dataset did not include a reliable individual identifier for matching each student's pretest and posttest scores, the analyses were conducted as independent-samples comparisons. First, scores from the Technical Baccalaureate group (experimental group) and the Science Baccalaureate group (control group) were compared separately at each testing moment. Second, pretest and posttest scores were compared within each group, using an independent-sample test. Therefore, the results describe differences between groups and testing moments, but they should not be interpreted as strict individual learning gains.

Table 1. Teaching Structure for the Experimental and Control Group

Week	Focus	Intervention group		Control group	
		Instructional activities	Evidence collected	Regular instructional activities	Evidence collected
1	Daily Routines vocabulary	The teacher introduced daily routines vocabulary using flashcards. Students looked at the pictures and tried to recall the English words before checking the correct answers. They then completed a worksheet to practice the new vocabulary.	Class notes, completed worksheets, teacher observation log.	Students attended their regular English classes following the institutional curriculum. The teacher introduced daily routine vocabulary using the course textbook and routine classroom activities. Students completed the corresponding textbook exercises	
2	Simple Present and Daily Routines	Students reviewed the vocabulary with flashcards by recalling the words from memory before checking the answers. They then completed a worksheet using the Simple Present to describe their daily routines.	Completed worksheets, participation records, teacher observation log.		Class notes, teacher observation log

Week	Focus	Intervention group		Control group	
		Instructional activities	Evidence collected	Regular instructional activities	Evidence collected
3	Vocabulary consolidation and posttest	Students reviewed the vocabulary learned during the previous lessons by recalling words with flashcards before checking the correct answers. They also completed short classroom activities to retrieve and consolidate the target vocabulary practiced during the intervention. Finally, they completed the posttest to evaluate vocabulary retention.	Completed review activities, posttest scores, teacher observation record.	Students continued learning and practicing the Simple Present through the course textbook, teacher explanations, and workbook activities. They practiced describing daily routines using regular classroom material. Unlike the experimental group, they did not participate in the structured flashcard-based retrieval and vocabulary. Students continued learning the Simple Present through the course textbook, teacher explanations, and workbook activities. They practiced describing daily routines using the regular classroom material.	Students' worksheets, participation records, observation log.

Source: own elaboration.

Note. The structure presented establishes the differences between the experimental group and the control group, which continued with regular instruction during the same period.

Data were analyzed by comparing pretest and posttest mean scores within each group and by comparing posttest performance between the experimental and control groups. Descriptive statistics, such as mean scores, standard deviations, and percentage differences, were used to summarize the results (Field, 2018). Because there was no valid identifier available to link pretest and posttest observations at the individual level, comparisons between the two time points were conducted as independent samples. Consequently, pre-post differences were estimated within each group using Welch's t-tests, supplemented with effect sizes,

Results

Table 2 compares the Science (control group) and Technical (experimental) tracks separately at each testing moment. In the pretest, most between-group differences were small and not statistically significant; however, the Technical Track already performed significantly better than the Science Track at the 2,000-word level. This baseline difference indicates that the groups were not fully equivalent before the intervention and should be taken into account when interpreting the posttest results. However, there was no statistically significant between-group difference in the total vocabulary score at baseline. The pattern changed in the posttest. At this point, the Technical Track scored higher than the Science Track across all vocabulary levels and in the total score,

with statistically significant differences in every comparison. The size of these differences was also larger than in the pretest, especially for the total score and for the 1,000-, 3,000-, 4,000-, and 5,000-word levels. Thus, the table suggests stronger post-test performance in the Technical Track group. However, the analysis compares groups at each moment rather than matched individual changes; these results should be considered as between-group differences over time, which means that there is no direct evidence of individual improvement.

Table 2. Descriptive Statistics and Between-Group Comparisons by Testing Moment

Moment	Vocabulary level	Science mean	Technical mean	d Difference	T	df	p	Cohen's d	d
Pretest	1000	14.24	15.00	-0.76	-0.274	40.756	0.785	-0.081	0.081
Pretest	2000	7.24	11.96	-4.72	-2.205	43.415	0.033*	-0.620	0.620
Pretest	3000	7.86	10.46	-2.60	-1.364	44.605	0.179	-0.387	0.387
Pretest	4000	7.10	7.65	-0.56	-0.322	44.018	0.749	-0.094	0.094
Pretest	5000	6.76	7.85	-1.08	-0.748	44.817	0.458	-0.213	0.213
Pretest	total	43.19	52.92	-9.73	-1.193	44.72	0.239	-0.339	0.339
Posttest	1000	9.92	17.67	-7.75	-3.561	41.96	0.001*	-1.028	1.028
Posttest	2000	7.38	12.17	-4.79	-2.475	42.673	0.017*	-0.714	0.714
Posttest	3000	7.42	13.00	-5.58	-3.44	35.592	0.001*	-0.993	0.993
Posttest	4000	5.21	10.83	-5.63	-3.338	31.5	0.002*	-0.964	0.964
Posttest	5000	5.13	11.71	-6.58	-3.914	37.334	0.000*	-1.130	1.130
Posttest	total	35.04	65.38	-30.33	-3.978	38.697	0.000*	-1.148	1.148

Source: own elaboration.

Note. Difference was calculated as Science Track Minus Technical Track. Negative values indicate higher scores in the Technical Track. Welch's independent-samples t test was used. Asterisk (*) means that it is a significant difference.

Table 3 presents a comparison between the pretest and posttest within each group. It was treated as a comparison of independent samples because no reliable individual identifier is available to pair the measurements. In First Grade Science (control group), no statistically significant differences were observed between the two time points, although scores tended to be lower on the posttest, especially at the 1,000-word level and for the total score. In First Grade Technical (experimental group), however, the means increased at all levels, although the only statistically significant difference was found at the 5,000-word level, with a moderate effect size.

Table 3. Independent Pretest–Posttest Comparisons Within Each Group

Group	Vocabulary level	Pretest mean	Posttest mean	Difference	t	df	p	Cohen's d	d
Science	Score_1000	14.24	9.92	-4.32	-1.721	32.9	0.095	-0.530	0.530
Science	Score_2000	7.24	7.38	0.14	0.080	42.0	0.937	0.024	0.024
Science	Score_3000	7.86	7.42	-0.44	-0.307	34.9	0.760	-0.094	0.094
Science	Score_4000	7.10	5.21	-1.89	-1.328	31.1	0.194	-0.411	0.411
Science	Score_5000	6.76	5.13	-1.64	-1.294	42.0	0.203	-0.387	0.387
Science	Score_total	43.19	35.04	-8.15	-1.237	39.2	0.224	-0.374	0.374

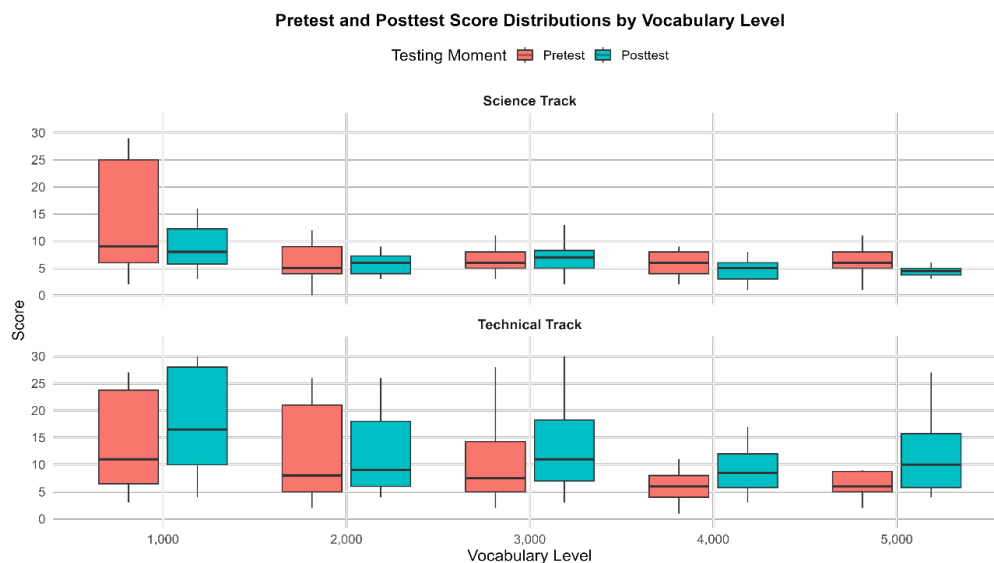
Group	Vocabulary level	Pretest mean	Posttest mean	Difference	t	df	p	Cohen's d	d
Technical	Score_1000	15.00	17.67	2.67	1.075	47.9	0.288	0.304	0.304
Technical	Score_2000	11.96	12.17	0.21	0.088	47.8	0.930	0.025	0.025
Technical	Score_3000	10.46	13.00	2.54	1.235	48.0	0.223	0.349	0.349
Technical	Score_4000	7.65	10.83	3.18	1.624	44.4	0.111	0.464	0.464
Technical	Score_5000	7.85	11.71	3.86	2.117	44.0	0.040	0.605	0.605
Technical	Score_total	52.92	65.38	12.45	1.381	47.8	0.174	0.391	0.391

Source: own elaboration.

Note. Difference was calculated as posttest minus pretest. Positive values indicate higher posttest scores. Welch's independent-samples t test was used because individual pretest–posttest pairing was not available.

Figure 1 provides a more detailed view of the distribution of scores by lexical level. In Science, the boxes show a general trend toward stability or a decrease in scores between the pretest and posttest, while in Technical, a more noticeable shift toward higher values is observed, particularly at the 4,000- and 5,000-word levels. This visual interpretation is consistent with the results shown in the previous table, although it should be interpreted with caution due to the lack of individual matching.

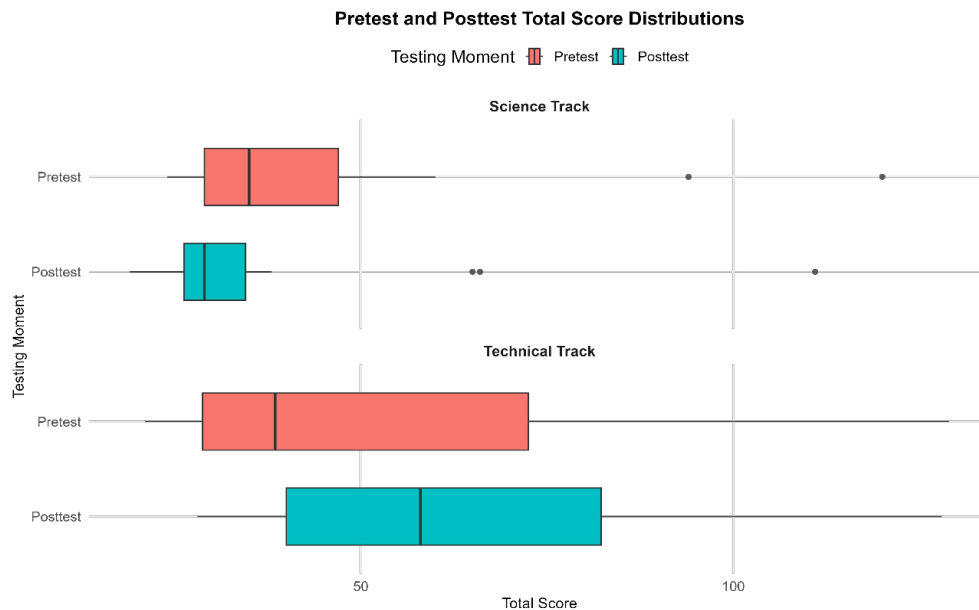
Figure 1. Pretest and Posttest Score Distributions by Vocabulary Level and Group



Source: own elaboration.

If we look at the second figure, we can see how the total scores vary between the two groups. The science group places the boxes in lower positions than before, while in the technical group, the subsequent scores tend to be higher. Based on this observation alone, we might begin to imagine that the technology group performed much better. However, the experimental design showed that other variables need to be controlled to provide a more comprehensive comparison between the two groups, one that considers the number of controlled elements.

Figure 2. Pretest and Posttest Total Score Distributions by Group



Source: own elaboration.

Discussion

The results showed that the science and technical specializations developed with different outcomes, which would initially suggest that in limited intervention processes like the one presented, results tend to vary. When the study began, the groups already showed a difference; this group advantage was certainly unexpected, but the technical specialization had progressed to the 2000-word level. It is important to note this fact because the intervention group started with an advantage, which could have influenced the final results. However, since this was a quasi-experiment in which we worked with intact groups of students, with the same teacher instructing both groups at different times, it was not possible to combine the students to homogenize the initial sample. Despite this, the intervention proceeded.

As we have seen in the results, the level achieved by the technical group is higher than that reached by the science group in all aspects, leading to a much higher total score for the technical group than for the science group. This situation leads to believe that repeated exposure, retrieval practice, and interaction with the productive context in which writing and speaking occur generate significant results, a finding also supported by authors such as Nation (2013) and Webb & Nation (2017). It is important to note that those in the technical group even significantly increased their level at the 5,000-word mark, demonstrating a much more consolidated receptive level than those who worked with the regular curriculum. Active retrieval, as suggested by authors such as Karpicke & Roediger (2008) and Roediger & Karpicke (2006), ultimately promotes retention more effectively than passive review, in which the student merely reads or listens without internalizing the meanings of the words.

As we have already pointed out, the two groups were not balanced from the beginning, so it is difficult to draw conclusions without pointing out this limitation. It is important to mention here that it was not possible to carry out a pre-post comparison with each individual because the tests were anonymous, in this sense the results were compared as coming from independent samples. In this way, the analysis limits the evaluation of rigor between gains obtained by each student. Differences between groups show a group-level effect, but do not do so at the individual level. This problem is important because the score was higher than in the pretest or posttest for science or technical, but aspects related to initial or final absenteeism were not controlled since the group of students approached the final exam period, which could possibly have affected their interest in the subject.

The graphs show that the group of students of the technical specialty tends to move the boxes towards the most ascending part in the posttest, while the non-intervened group, which, in principle, should remain as it is at the beginning, paradoxically tends to be located at the lower values. Explaining this situation is difficult as the author of this work because an external evaluator could think that there was manipulation of data when it was not the case. In any case, the group composition can be blamed since in the science group at the beginning there were only 21 students and in the posttest 24 responded, while in the technical group at the beginning there were 26 and then there were only 24 students. In this sense, the evaluation of the technical group retained the same students, although with absences, while the science group included students who did not take the pretest.

These patterns indicate differences between the groups at the posttest, however, they do not establish that the intervention caused the observed changes, particularly given the baseline differences between the groups. Another possible explanation for this situation is that participants may have been aware that they were being observed and that their results were compared with the other group. According to McCambridge et al. (2014), it could cause a Hawthorne effect because they knew that they were observed and evaluated, which could increase their effort and participation.

Previous research has similarly shown that retrieval practice, repeated vocabulary practice, and contextualized learning can support vocabulary development. Previous studies suggested that flashcards and repeated vocabulary practice can support learning (Nakata, 2011; Webb, 2007; Karpicke & Roediger, 2008; Komachali & Khodareza, 2012). In the present study, the Technical Track evaluation showed higher posttest scores than the Science Track at all vocabulary levels and in the total score. This result may be related to the repeated work with vocabulary, recall activities, and review included in the intervention. However, the technical group already showed a statistically significant advantage at the 2,000-word level in the pretest, and the data do not allow individual pretest-posttest matching. Therefore, the results should be interpreted as better posttest performance in the experimental group, but not as conclusive evidence that the posttest differences were caused exclusively by the intervention. This baseline difference limits any causal interpretation and suggests that pre-existing group differences may have contributed to the posttest results.

Overall, the study suggests that vocabulary instruction based on repetition, retrieval, and contextualized activities may support receptive vocabulary development, especially when classroom conditions favor engagement with the target words. In any case, future vocabulary studies could include some other identification method without using names, or they could divide the courses into two groups of equivalent levels. They could also conduct the experiment with a second test after a considerable time to measure retention of the new vocabulary over time. This could help make this line of research more precise by controlling individual progress over time.

Pedagogical Implications and Limitations

The implications of the findings presented regarding the evidence of retrieval, repetition, and contextualized activity refer to the simultaneous combination of these three, not their separate use. This integration is important because it points toward a more systematic method suggested by the literature presented.

Another implication relates to the teaching practice of those who teach EFL, especially in places where regular exposure to the language is very limited. It is important to consider that in many Amazonian areas, such as Limón Indanza, the formalization of secondary education must contend with complex migration processes, a representative parental absenteeism, students from rural areas with limited cultural capital, and linguistic interference, particularly in the case of students from the Shuar ethnic group.

Teachers should see the combination of retrieval, repetition, and context as an opportunity to make vocabulary learning a more engaging process, enabling students to become more familiar with EFL. Furthermore, it is important to mention some sampling limitations, such as the small sample size and the limited time available, which is only three hours per week, resulting in a total intervention of nine hours. Future research should address this issue, but also consider controlling for other psychological variables, using an identifier that does not necessarily have to be the student's name, conducting the research at a time that avoids stress close to exams, and above all, measuring what happens in the long term, after a term or the academic year, to determine if retention was significant.

Another limitation would be controlling external factors such as students' exposure to the foreign language outside the classroom, such as their study habits. These variables influence the results, as the level of familiarity with the language is an important factor. To address this, random assignment of students to groups could be considered, when possible, within the classroom itself, while they are carrying out tasks.

In conclusion, retrieval, repetition, and activities within a context generated evidence regarding vocabulary teaching and learning. Interaction and activities that engage students in learning new words provide support for the more authentic development of lexical skills in the classroom environment. The technical specialization group showed that, after intervention, they were able to

achieve better results than the science specialization group, especially at the 4,000 and 5,000 word levels of the Webb test. However, it is important to remember that the two groups started from different baselines, so they were not entirely equivalent. Therefore, we interpret the results with caution and do not generalize them. Despite these limitations, this work generated contextualized evidence and could be a precedent for other studies dealing with their own specific contextual conditions.

References

- Abad Célleri, M. P., Argudo Serrano, J. C., Fajardo Dack, T. M., & Cabrera Tenecela, H. P. (2019). English proficiency and learner individual differences: A study of pre-service EFL student-teachers. *Maskana*, 10(1), 1–9. <https://doi.org/10.18537/mskn.10.01.01>
- Cabrera-Tenecela, P. (2023). New organization of research designs. *South American Research Journal*, 3(1), 37–51. <https://doi.org/10.5281/zenodo.8050508>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Field, A. (2018). *Discovering statistics using IBM SPSS Statistics*. SAGE Publications.
- Godwin-Jones, R. (2020). Building the porous classroom: An expanded model for blended language learning. *Language Learning & Technology*, 24(3), 1–18. <https://doi.org/10.64152/10125/4473>
- Hacking, J. F., & Tschirner, E. (2017). The contribution of vocabulary knowledge to reading proficiency: The case of college Russian. *Foreign Language Annals*, 50(3), 500–518. <https://doi.org/10.1111/flan.12282>
- Joe, A. (1995). Text-based tasks and incidental vocabulary learning. *Second Language Research*, 11(2), 149–158. <https://doi.org/10.1177/026765839501100206>
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. <https://doi.org/10.1126/science.1152408>
- Kline, T. J. B. (2017). Sample issues, methodological implications, and best practices. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, 49(2), 71–77. <https://doi.org/10.1037/cbs0000054>
- Komachali, M. E., & Khodareza, M. (2012). The effect of using vocabulary flash card on Iranian pre-university students' vocabulary knowledge. *International Education Studies*, 5(3), 134–147. <https://doi.org/10.5539/ies.v5n3p134>

- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399–436. <https://doi.org/10.1111/j.0023-8333.2004.00260.x>
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26. <https://doi.org/10.1093/applin/22.1.1>
- Li, J., Zhang, E.-H., Zhang, H., Gou, H., & Cao, H.-W. (2024). Retrieval practice plus feedback benefits a third language vocabulary learning. *International Journal of Multilingualism*, 21(2), 1014–1033. <https://doi.org/10.1080/14790718.2022.2102172>
- López, P. L. (2004). Población, muestra y muestreo. *Punto Cero*, 09(08), 69–74.
- McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, 67(3), 267–277. <https://doi.org/10.1016/j.jclinepi.2013.08.015>
- Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, 24(1), 17–38. <https://doi.org/10.1080/09588221.2010.520673>
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge University Press.
- Qu, K., Liu, T., Qiao, Y., & Wang, P. (2024). The facilitative effect of the keyword mnemonic on L2 vocabulary retrieval practice. *Heliyon*, 10(3). <https://doi.org/10.1016/j.heliyon.2024.e25212>
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- Ushioda, E. (2021). Doing complexity research in the language classroom: A commentary. In R. J. Sampson & R. S. Pinner (Eds.). *Complexity perspectives on researching language learner and teacher psychology* (pp. 269–283). Multilingual Matters. <https://doi.org/10.21832/9781788923569-017>
- Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(1), 33–52. <https://doi.org/10.1017/S0272263105050023>
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46–65. <https://doi.org/10.1093/applin/aml048>
- Webb, S., & Chang, A. C.-S. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, 19(6), 667–686. <https://doi.org/10.1177/1362168814559800>
- Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL - International Journal of Applied Linguistics*, 168(1), 34–70. <https://doi.org/10.1075/itl.168.1.02web>

Authors

Narcisa de Jesús Gómez León. Master's student in Teaching English as a Foreign Language at Universidad Católica de Cuenca. Engineer in Tourism and Hotel Business Administration. Currently works as an English teacher at the Ministry of Education of Ecuador.

Juanita Argudo-Serrano. Associate Professor at the English teaching training program at Universidad de Cuenca, and a teacher at the Teaching English as a Foreign Language Master's Program at Universidad Católica de Cuenca, both in Cuenca, Ecuador. Juanita is a researcher at Universidad de Cuenca and collaborates on several projects at different universities in Cuenca. She is also a researcher in the EFL Teaching Education research group and the Director of the Education Department at Universidad de Cuenca. Her teaching and research areas of interest are in-service and pre-service teachers' development, language evaluation, content and language-integrated learning for pre-service and in-service teachers, expressive writing to relieve stress, and critical language pedagogy.

Declaration

Conflict of interest

We have no conflicts of interest to declare.

Funding

No financial support from external parties for this article.

Note

This article is the result of the author's degree work as part of the master's program in Teaching English as a Foreign Language at Universidad Católica de Cuenca.