The Impact of Gamification on Vocabulary Acquisition

A Comparative Study of Two Student Response Systems in Enhancing Vocabulary Acquisition

Cindy Xerri
The Impact of Gamification on Vocabulary Acquisition

A Comparative Study of Two Student Response Systems in Enhancing Vocabulary Acquisition

Cindy Xerri

Abstract

As digitalization continues to advance, digital tools have become omnipresent in the classroom, and Student Response Systems (SRSs) have emerged as a valuable tool for teachers. However, the added value and efficacy of integrating gamified SRSs, compared to non-gamified SRSs, in the classroom has yet to be fully explored. Furthermore, little attention has been paid to non-gamified SRSs. The present study examines how a gamified learning environment in SRSs impacts vocabulary acquisition compared to a non-gamified learning environment. To analyze this effect, a quantitative study was designed using statistical analyses such as the Shapiro-Wilk test, the Friedman test and the Wilcoxon signed-rank test. SRSs, namely Gimkit and Socrative, were selected to measure vocabulary acquisition.

A total of 61 high school students studying English were enrolled in this study. Eighteen target words have been picked from two different vocabulary books designed to help students reach the C1 level in English. The selected words have all been identified as adjectives. The students were split into two class-based groups, and each was assigned an SRS to use for two 20-minute study sessions, happening at an interval of three days. To measure the effect of the two SRSs on vocabulary acquisition, three tests were conducted over three weeks. The students took a pre-test, an immediate post-test after the second study session, and a final delayed post-test two weeks later.

The statistical analysis and the results of the three tests measuring vocabulary acquisition revealed a statistically significant improvement in vocabulary test scores for both groups in the short- and long-term. However, no statistical significance was shown for the test scores between the two groups: gamified and non-gamified learning environments. Hence, a gamified learning environment using SRSs was shown to have no statistical significance on vocabulary learning and retention compared to a non-gamified one. While digital tools are widely embraced, this study suggests that gamified SRSs, compared to non-gamified SRSs, may not offer a significant advantage in terms of vocabulary acquisition and retention. It is, therefore, important for the teachers to get to know their students and find the type of SRSs that suit their learning style best. Further studies could investigate the impact of gamified and non-gamified SRS on the students’ vocabulary acquisition and motivation over a school year to understand the effects better.

Keywords

EFL, vocabulary acquisition, SRS, gamified learning environment, adjectives.
1. Introduction

Within the domain of language acquisition, vocabulary serves as a fundamental component. Indeed, learning vocabulary is vital to expressing thoughts, emotions, and ideas as freely as possible. In particular, adjectives are essential to add depth and nuance to one’s vocabulary while developing fluency and proficiency. Fostering vocabulary acquisition in education is primordial to help students develop their language competence and can be done using different tools and methods. Meanwhile, the technological development and digitalization of our society have led to some changes in our ways of teaching and learning. In the Swedish context, this shift is supported by Skolverket (2023), the Swedish National Agency for Education, which encourages the use of technology in the classroom. During the past years, the methods and tools for practicing vocabulary have changed, and the traditional paper-based approach has been joined by an increasing number of digital tools and platforms to facilitate learning. This, therefore, raises some questions about the efficacy of those new digital methods compared to the previous paper-based methods.

A noteworthy digital tool in this context is Socrative, a Student Response System (SRS) that allows students to train on any device with a web browser (Mork, 2014). Socrative offers students instant feedback as they complete questions such as true or false, multiple-choice questions, or even short answers (Balta & Tzafilkou, 2019). Moreover, Socrative also enables teachers to monitor the students’ responses in real-time, facilitating the identification of areas necessitating further reviews. Socrative has even been classified as one of the six technological innovations that transformed how English is taught, making it a great asset in the classroom (Jana & Iveta, 2019). A platform different from Socrative is Gimkit, which enables the students to train in a gamified way, where correct answers allow them to earn points that they can later use as rewards. Therefore, these two platforms have different approaches towards learning, Socrative offering a non-gamified approach and Gimkit offering a gamified one, while both enable the students to exercise the same kind of questions.

Employing SRSs and whether the motivation generated by impact vocabulary acquisition have received some attention in Saari and Varjonen (2021) who analyzed how using SRSs and the motivation generated by these activities impact vocabulary acquisition. The findings from this study identified that the students considered the use of SRSs to be positive for their learning, and it also revealed the positive impact of these two gamified SRSs on the students’ motivation. However, this study focused solely on gamified SRSs. Amidst the substantial number of studies on gamified Student Response Systems (SRS) in education and their use in different subjects (Pede, 2017), most studies focus on the gamified aspect of these tools and their impact on engagement and collaboration (Rofiah & Waluyo, 2020; Saeed Alharbi & Meccawy, 2020). As Vanderbruyse et al. (2013) established, more studies need to be done to evaluate the added value and efficacy of game-based learning. Furthermore, very little space has been conceded towards non-gamified SRSs, such as Socrative. Several studies (Adnyani et al., 2020; Licorish et al., 2018), conducted mainly on gamified SRSs, revealed positive perceptions of digital tools
from both learners’ and teachers’ perspectives. However, the effect of gamified and non-gamified SRS on vocabulary acquisition is yet to be compared.

While previous studies have predominantly explored gamified SRSs and their positive effect on motivation and engagement, there is a notable lack of research comparing the efficacy and impact of gamified and non-gamified SRS on vocabulary acquisition. The present study will investigate the impact of gamification on vocabulary acquisition using two types of SRS: Gimkit, with a gamified learning environment, and Socrative, with a non-gamified learning environment. This study will focus on vocabulary acquisition of adjectives. The teaching approach can profoundly impact the students’ learning path; hence, choosing different ways to encourage vocabulary learning is crucial to provide the students with all the tools possible to succeed in their learning journey. The study aims to shed light on the strengths and weaknesses of each approach and consider the implications of their use in the classroom.

This present study will try to answer the following question:

- How does a gamified learning environment in SRS impact vocabulary acquisition?

The research question aims to analyze whether the gamified nature of the learning environment, combined with the time dimension of short- and long-term retention, plays a pivotal role in influencing students’ vocabulary acquisition. Formulating hypotheses sets the outline and framework for interpreting the results gathered in this study. In the present study, two main effect hypotheses are being tested: whether the gamified environment in SRSs has a significant effect on vocabulary acquisition compared to a non-gamified environment and whether time (retention short-term and long-term) has a significant effect on vocabulary acquisition. Moreover, it also becomes relevant for this study to analyze the interaction effect between the two main effects of environment setting and time. Hence, the interaction hypothesis tests whether a significant interaction effect exists between the environment setting and time on vocabulary acquisition.

2. Literature Review

To analyze and compare the impact of gamified and non-gamified SRS on vocabulary acquisition, it is essential first to get a better understanding of vocabulary learning and what it entails, as well as an understanding of how memory works. This literature review will then address the place technology can have in education, as well as Skolverket’s guidelines on the subject as this study takes place in the Swedish context.

2.1 Vocabulary Learning

Learning is considered to be a stable change in a person’s knowledge and competence (Spector & Kim, 2014). When learning a new language, it is essential for learners to develop their vocabulary to express themselves without limitations. A more extensive vocabulary is often linked to a higher proficiency. Hence, learning vocabulary is essential to mastering a language (Schmitt, 2010). Consequently, vocabulary knowledge positively
impacts the different competencies related to knowing a language: reading, listening, writing, and speaking. Vocabulary knowledge has, therefore, an influence on both productive and receptive skills and is an influential parameter in language learning. Karami & Bowles (2019) emphasized the importance of vocabulary in language production, echoing Nation’s (2001) sentiments who asserted that “vocabulary and language have bidirectional effects in such a way that knowledge in vocabulary can result in the use of language and the use of language can result in the knowledge of vocabulary” (as cited in Karami & Bowles, 2019, p. 26).

The learning process can occur in diverse contexts and have bidirectional effects on the learner’s capacities. Within this continuous learning process, two distinct kinds of learning emerge: incidental and intentional. Kerka (2000) defines incidental learning as “unintentional or unplanned learning that results from other activities” (p. 3). Huckin & Coady (1999) further elaborate that incidental word acquisition necessitates the learner to encounter the word several times in different contexts, emphasizing the longitudinal nature of vocabulary learning. However, there is no consensus in research concerning the exact number of expositions required to acquire the target word (Huckin & Coady, 1999). In opposition to incidental learning, intentional learning is defined by Hulstijn (2013) as “a deliberate attempt to commit factual information to memory, often including the use of rehearsal techniques, like preparing for a test in school or learning a song by heart” (p. 1). It is also important to note that vocabulary development and acquisition of new words is an incremental process in which the learners gradually acquire some word knowledge over time by encountering the word in different forms and contexts (Schmitt, 2010). With increased exposure to vocabulary, the learner will consolidate the knowledge acquired about the word and learn new meanings about the said vocabulary item (Schmitt, 2010). Vocabulary learning, therefore, happens in different conditions and is a longitudinal process requiring several expositions to the target words to be learned and for the learners to be able to use them.

2.1.1 Knowing a Word

It is essential to establish what knowing a word entails. Indeed, as Nation (2013) explains, words are not isolated units and are instead part of a system where units are interconnected. The first aspect that Nation (2013) highlights is the learning burden. Depending on the degree to which the learner is familiar with the word’s pattern and knowledge, the less of a burden it is for the learner as they can be (Nation, 2013). A distinction can also be drawn between knowing a word receptively and knowing a word productively. Indeed, Nation (2013) explains that knowing a word receptively means that the form of the word is perceived while reading or listening, and the learners then identify its meaning, whereas knowing a word productively means using the word to communicate when writing or speaking and requires the word to be used in its appropriate spoken and written form. Table 1, extracted from Nation’s work, illustrates the multifaceted aspects of knowing a word:
In the present study, the students will exercise receptive and productive vocabulary; it is therefore important to understand what it entails. As Table 1 displays, knowing a word receptively entails knowing how the word sounds as well as being able to recognize the word by the way it is spelled, and knowing in which kind of context/collocations it is used, and what it means (Nation, 2013, pp. 9-10). As visible in Table 1, knowing a word receptively also entails knowing the kind of words the target word occurs with, such as when and how often the learner can be expected to meet that word. Knowing a word receptively is not only limited to understanding it but also includes recognizing it orally and in its written form. In the continuity of receptive knowledge, knowing a word productively entails knowing the pronunciation, spelling, and collocations, as well as knowing when and how often to use it. Productive knowledge of words also encompasses knowing what kind of grammatical pattern the word should be used, as visible in the above table (Table 1). Learners tend to first master the word from a receptive point of view before using it productively, as learning a word productively may require more exposition and rehearsal (Laufer, 2005). Griffin (1992) pointed out that productive

<table>
<thead>
<tr>
<th>Form</th>
<th>Spoken</th>
<th>Receptive</th>
<th>What does the word sound like?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Productive</td>
<td>How is the word pronounced?</td>
</tr>
<tr>
<td>Written</td>
<td>Receptive</td>
<td>What does the word look like?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Productive</td>
<td>How is the word written and spelled?</td>
<td></td>
</tr>
<tr>
<td>Meaning</td>
<td>Form and meaning</td>
<td>Receptive</td>
<td>What meaning does this word form signal?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Productive</td>
<td>What word form can be used to express this meaning?</td>
</tr>
<tr>
<td>Associations</td>
<td>Receptive</td>
<td>What other words does this make us think of?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Productive</td>
<td>What other words could we use instead of this one?</td>
<td></td>
</tr>
<tr>
<td>Use</td>
<td>Grammatical functions</td>
<td>Receptive</td>
<td>In what patterns does the word occur?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Productive</td>
<td>In what patterns must we use the word?</td>
</tr>
<tr>
<td>Collocations</td>
<td>Receptive</td>
<td>What words or types of words occur with this one?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Productive</td>
<td>What words or types of words must we use with this one?</td>
<td></td>
</tr>
<tr>
<td>Constraints on use (register, frequency …)</td>
<td>Receptive</td>
<td>Where, when, and how often would we expect to meet this word?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Productive</td>
<td>Where, when, and how often can we use this word?</td>
<td></td>
</tr>
</tbody>
</table>
vocabulary learning seems more problematic for learners. Hence, learning vocabulary productively could require more time and effort than learning vocabulary receptively.

2.1.2 Role and Types of Adjectives

Adjectives are essential to a language and enable the speaker to elevate their communication and share their ideas and visions precisely. As Fowler (2016) explains, for a century, the word *adjective* functioned as an adjective before its adoption as a noun used to denote one of the parts of speech. Greenbaum (1996) defines adjectives the following way: “An adjective is a word such as *wise* that typically can premodify a noun such as *decision (a wise decision)* and function as subject predicative after a copular verb such as *be* or *seem (The decision is/seems wise)*. When used as the premodifier of a noun, the adjective is attributive; when used as subject predicative, it is predicative” (p. 615). Adjectives are, therefore, modifiers enabling the characterization or description of nouns or pronouns. Adjectives can, therefore, help provide more information about the noun for which they qualify.

Fowler (2016) further explains that most adjectives possess the flexibility to be used in two distinct positions: either before the noun, in an attributive position and will often denote characteristics of the word (example: a white cat), or after the word, usually separated by a verb of state in predicative position (example: The cat is white). Moreover, Some adjectives are restricted to having a predicative or an attributive position, such as *afraid*, which can only have a predicative position (Fowler, 2016).

2.2 Memory and Learning

The word *memory* is primarily used as a general word to describe cognitive function; however, it is essential to distinguish between the different kinds of memory. There are two categories of memory: short-term memory and long-term memory. A distinction can be drawn between the sensory memory linked with the senses, the working memory that handles the information during a short period, and the long-term memory. The long-term memory is then self-divided into two parts, the explicit and implicit memory. Explicit memory contains episodic memory, which usually includes some special events that happened throughout the individual’s life, and semantic memory contains the individual’s knowledge of the world (Roediger et al., 2017). Semantic memory contains information about words, such as their meaning; this area can also know about different concepts or historical events and when they took place (Roediger et al., 2017). This area of memory is broad and includes much information of various kinds.

In the context of vocabulary acquisition, semantic memory is of interest. Even though long-term memory possesses an unlimited capacity, the items can be forgotten, as Hermann Ebbinghaus (2013) shows with the concept of the *forgetting curve*. Indeed, the Ebbinghaus curve (2013) shows that the learner forgets quickly after learning; however, the forgetting rate gradually diminishes over time. Murre & Dros (2015), who replicated the study of Ebbinghaus in the Dutch context, obtained similar results to Ebbinghaus (2013), which are the following: Learners tend to forget half of the newly acquired information within the first hours or days after the first learning session unless the new
information is studied again (Murre & Dros, 2015). It had also been shown that the more an item has been studied, the slower the individual will forget about it (Loftus, 1985). There seems to be an interest in encouraging the student to overlearn certain information, which could generate better retention. Repetitions, therefore, seem to play a crucial role in learning. Spaced repetition, characterized by short study sessions spread over time, has been highlighted by its positive effects compared to massed practice, which tends to result in shorter-term retention of the information (Kornell, 2009). Learning language is a cumulative process that builds up on previous knowledge; it is therefore essential that the knowledge acquired by the learners remains to improve their language competence (Seibert Hanson & Brown, 2020). Several studies support that spaced reception positively affects learning (Ellis, 1995); (Bloom & Shuell, 1981). These positive results obtained by spaced repetition, therefore, raise attention to the fact that spaced repetition should be included in teaching more often. Structuring instruction to revisit studied items based on the students’ needs regularly would be beneficial, and SRS can be a valuable tool in this process.

2.3 Technology in Education

2.3.1 SRS in Education

SRSs, also called clickers (Patterson et al., 2010), have become increasingly common in the classroom in previous years. SRSs can have different uses in the classroom; they can, for example, be used to measure the students’ opinion on matters, encourage interactions between students or between students and teachers, and can even be used for assessment purposes (Cerqueiro & Harrison, 2019). SRS can, therefore, be used in different ways depending on the activity and the desired learning outcomes. SRSs are also perfectly adaptable to large groups, such as those found in the classroom (Faya Cerqueiro & Martín-Macho Harrison, 2019). These systems enable teachers to be aware of the students’ learning path and adjust their teaching methods accordingly.

SRSs also typically provide immediate formative feedback, enabling students to learn and memorize as they complete the questions. Shapiro & Gordon (2012) conducted a study on the effect of SRS questions on the students’ memory. The participants involved in this study were studying psychology. This study revealed that the students who trained on SRSs tended to score higher in delayed tests, demonstrating the efficacy of SRSs. This consequence has also been pointed out in the study as being related to the testing effect, which means that the learner’s long-term memory is increased when they are asked to retrieve the information during the learning period. This shows how SRS can efficiently be used for learning to enhance the students’ memory of the target information or element.

Patterson et al. (2010), who studied the use of SRSs within a large classroom of nursing students, also noticed that using SRSs in the classroom results in positive outcomes as using SRSs increased the students’ perceived engagement and interaction within the classroom, enabling them to take a more active part in the lesson. These findings are similar to those of Sénécal et al. (2022). Indeed, Sénécal et al. (2022) sought to analyze learners’ and teachers’ perceptions of SRSs based on both the students’ learning and self-assessment and the engagement generated by using SRSs in the classroom and their interactivity. They focused on the impact of SRSs on vocabulary learning, revealing the
students’ positive attitudes toward using SRSs in the classroom compared to those who trained without using SRSs. This study showed that students tend to have a positive attitude towards SRSs and consider them to impact their engagement in the lesson positively. In contrast, teachers tend to have a neutral or negative attitude concerning using SRSs (Sénécal et al., 2022).

2.3.2 Active Learning and Testing Effect

Engaging in quizzes using SRSs enables the students to engage actively with the material to be learned. This allows the learner to actively engage instead of passively studying it, leading to better retention, as confirmed by Roediger & Butler’s (2011) study, which showed that enabling the learners to read the information and test themselves yields better results than reading the target information several times. Hence, allowing the learners to test their knowledge is critical to provide them with the best learning tools. Using SRSs enables the student to actively engage with the target information, leading to better retention.

Campbell and Mayer (2009) conducted a laboratory experiment on the question’s effect on the students’ cognitive retention when using SRSs. This study highlighted the fact that students tend to learn better when required to answer questions during lectures and were provided with formative feedback using SRSs, compared to when only presented with the question, like in a lecture. Therefore, SRSs enable students to test their knowledge, assisting them in participating and engaging actively. Moreover, SRSs allows students to get direct feedback after answering a question.

2.3.3 The Impact of Immediate Feedback

Feedback plays a critical role in the learning process and helps the learner gauge their progress and cultivate their self-regulated learning competencies. Indeed, according to the self-regulated learning model of Winne & Hadwin (1998), immediate feedback can help the learner to become aware of their progression and know if more efforts need to be invested, which in turn can help to develop the learner’s self-regulated learning competencies (Butler & Winne, 1995). Those results are also supported by Schunk (2003), who explains that providing learners with immediate feedback enables them to see their learning progress and have better knowledge of the learning goals, positively affecting their motivation. It has also been shown that a higher level of self-regulated learning competence results in higher motivation in the learners, resulting in better learning results (Zimmerman, 2002).

When taking tests and examinations, students may typically get the correction and receive feedback several days or weeks after completion. Lu et al. (2023) studied the effect of both immediate and delayed feedback in language learning exercises. They revealed that students who received feedback, both delayed and immediate, generally tended to score higher in the post-assignment. Moreover, the study showed that students who received immediate feedback had significantly higher grades than students who received delayed feedback at the end of the assignment. This highlights the positive effects of immediate
feedback compared to delayed feedback, which allows the students to integrate the target information better, leading to better learning outcomes.

2.3.4 Gamified & Game-based Language Learning

An alternative to word lists and other traditional learning methods can be gamified learning tools. Deterding et al. (2011) define gamification as “The use of game design elements in a non-game context” (p. 10). Some gamification elements in SRSs can for example be avatars, points, competition and time limit (Tan & Saucerman, 2017). An example of a gamified SRS is Quizlet. Zhou (2016) investigated how the use of Quizlet impacts the students’ experience when learning vocabulary; this study took place in the Chinese context and involved forty-six students who studied Chinese as a second language and did not have any prior notion of that language before engaging in the course. This study revealed that most students had a positive attitude concerning using Quizlet to study vocabulary and considered it useful. Zhou (2016) explained that some students even used Quizlet outside the classroom for the same purpose. She also pointed out that including Quizlet in the classroom develops extrinsic motivation in the students as many strive to get good results in competitions and have superior results compared to their coursemates (Zhou, 2016). The competitive elements within this platform should not be overlooked as they can strongly impact motivation and engagement with the task.

Rasti-Behbahani & Shahbazi (2022) studied the effect of digital games-based vocabulary learning on 124 Persian speakers who were studying English. The learners were divided into two groups; one group practiced on a digital game, whereas the other group practiced more traditionally by completing exercises such as fill-in-the-blanks and True or False and pairing a word with their definition. It was found that the students training on the game acquired both productive and receptive knowledge of the target vocabulary more efficiently than the control group who practiced using traditional methods (Rasti-Behbahani & Shahbazi, 2022). Moreover, they noticed that the students who practiced in a gamified manner improved their ability to recognize and understand the relationship between the form and its meaning, significantly better than the other students using traditional vocabulary learning methods (Rasti-Behbahani & Shahbazi, 2022).

Vandercruyssse et al. (2013) studied the impact of a competitive environment on students learning and perception. The eighty-three students involved in the study worked in a game-based learning environment to improve their business English conversational skills. This study revealed that students who trained in a competitive setting reported a higher perceived competence, a larger appetite for the task, and a higher level of enjoyment even though they still considered the activity a learning activity.

However, an element that might negatively impact the students’ learning is their tendency to assign less importance to the target elements due to the gamified nature of learning. Indeed, some students may invest less energy into learning when the activity is presented as a game (Salomon, 1984). This highlights the importance of presenting the learning goal of the activities to the students to demonstrate how those activities contribute to their learning.
2.3 Skolverket’s Recommendations

2.3.1 Skolverket and Vocabulary

In accordance with the guidelines from Skolverket (2021), the Swedish National Agency for Education, students must be afforded the opportunity to develop a comprehensive communicative competence. Skolverket (2021) then further elaborates that this encompasses both the comprehension of spoken language and texts, as well as expressing themselves and interacting with others in speech and writing. The students are also expected to be able to adapt their language depending on the situation, the purpose, and their interlocutor (Skolverket, 2021). Therefore, the students are expected to be able to produce, as well as understand, a wide range of vocabulary depending on the situation. Skolverket (2021) then explains that the students should be given chances to enhance their language accuracy in both spoken and written communication while fostering the capability to express thoughts with nuance and complexity. Possessing a wide range of vocabulary entails also knowing a large number of adjectives to express oneself in many ways, justifying the choice of this study to focus on vocabulary and, in particular, adjectives.

2.3.2 Skolverket and Digital Tools

Digital tools are now more and more present in education. Skolverket (2018) explains that the digital environment has become more dynamic in previous decades, and teachers and students use various tools and learning platforms. Today, the use of those tools is an important aspect of teaching. Skolverket (2018) explains that digitalization has changed the traditional teaching approach and how students appropriate knowledge and communication. Traditional teaching can now be complemented by web-based teaching, enabling teachers to have better control over how and what they want to teach (Skolverket, 2018).

Furthermore, Skolverket (2018) also argues that it is essential for teachers to know how digital tools can be used in teaching and be aware of the possibilities and limitations associated with these tools. Skolverket (2023) explains that providing students with different tools to explore and learn is essential. Indeed, this can help students find ways to learn that best suits them. Skolverket (2023) adds that the teacher is responsible for structuring and planning the teaching depending on what best suits the students’ needs. Therefore, this experiment will greatly help in learning more about the impact of SRS on vocabulary acquisition.

3. Methodology

The present study aimed to identify how a gamified learning environment in SRS impacts vocabulary acquisition. Two main effect hypotheses were to be tested: first, whether the gamified setting in SRS has a significant effect on vocabulary acquisition compared to a non-gamified setting, and second, whether time (retention short-term and long-term) has a significant effect on vocabulary learning. Moreover, the interaction hypothesis tests
whether a significant interaction effect exists between the setting and time on vocabulary acquisition. To do so, a quantitative and statistical analysis was conducted.

3.1 Study Design

Statistical analysis allows researchers to determine whether the means of multiple groups are equal or if there is a statistical significance among them, drawing conclusions about the relationships between variables and providing insights into the overall data patterns. Some models may help to analyze groups in relation to the variation within the groups, helping to determine whether any differences observed are due to random variation or if they are likely to result from some systematic effect. Choosing the appropriate statistical methods depends on various factors, including the nature of the data, the research question being addressed, and the specific assumptions of each statistical test.

In this setting, statistical analyses were used to determine if changes in test scores (the dependent variable) resulted from the interaction between the type of vocabulary training method (the between-subjects factor; gamified and non-gamified learning environment) and time (the within-subjects factor; consisting of different time points). To analyze the interaction between these two factors, a follow-up test using post hoc tests can be performed to determine whether changes in test scores can be solely attributed to one of the factors or influenced by a combination of the two.

3.1.1 Statistical Model

Based on the research question and the hypothesis originating from it, the following variables and factors needed to be controlled for:

Dependent variable: test score

Independent variables:
- Within-subjects factor: time
- Between-subjects factor: learning environment

Assumptions were considered and controlled based on the selected statistical model to analyze the data. Choosing the appropriate statistical analysis method depended on various factors, such as the nature of the data, the research question being addressed, and the specific assumptions associated with each statistical test. Considering these assumptions was imperative when constructing the study and subsequent tests.

3.2 Participants

The participants came from one of two classes in a school in metropolitan Stockholm, were all enrolled in the same educational program focused on economics and were studying English 6. Mackey and Gass (2021) explain that using preexisting classes can be “the most ecological sound setting for research” (p. 265). This choice considered the possibility that using two digital tools simultaneously in the classroom might distract some students and remove their attention from vocabulary training. Additionally, the
random assignment of each SRS to each class ensured that the learners’ profiles did not impact the effect of each tool on their learning outcomes.

### 3.3 Procedure and Material

Vocabulary tests were designed to determine the vocabulary knowledge of the students. The construction of these tests and the words used will be overviewed in future sections 3.4 and 3.3.2, respectively. These tests provide the test scores at each assessment point, which are used for subsequent analysis. The present study contains three assessment moments: a pretest before any session, to get the initial skill level; followed by an immediate post-test, to assess vocabulary knowledge after the final study session using the SRS; followed by a final delayed post-test. Further information concerning the reasoning behind the time spans between these assessments will be presented in the upcoming section 3.3.5.

To analyze the impact of the learning environment in this study, two test groups were created: one exposed to a gamified learning environment and the other to a non-gamified learning environment. Each group participated in two training sessions using the SRS attributed to their group. The chosen SRS platforms for this study were Socrative and Gimkit. The primary distinction between the two lies in the fact that Gimkit has a gamified approach to learning, whereas Socrative functions as a non-gamified SRS. The students have already been exposed to Socrative and Gimkit in an educational context and are familiar with how these software work. The characteristics of each SRS will be presented in the upcoming sections 3.3.3, 3.3.7, and 3.3.8.

Hence, the representation of the experiment’s setup became as follows.

**Table 2. Representation of the experiment’s setup.**

<table>
<thead>
<tr>
<th>Lesson</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson 1</td>
<td>Study agreement (cf. Appendix A)</td>
</tr>
<tr>
<td></td>
<td>Questionnaire (cf. Appendix B)</td>
</tr>
<tr>
<td></td>
<td>Pre-test</td>
</tr>
<tr>
<td></td>
<td>Learning activities (cf. Appendix C)</td>
</tr>
<tr>
<td></td>
<td>Study session 1 on SRS</td>
</tr>
<tr>
<td>Lesson 2 (2 days after lesson 1)</td>
<td>Study session 2 on SRS</td>
</tr>
<tr>
<td></td>
<td>Immediate post-test</td>
</tr>
<tr>
<td>Lesson 3 (2 weeks later)</td>
<td>Delayed post-test</td>
</tr>
</tbody>
</table>

As visible in Table 2, during the first lesson, the students were invited to sign the study agreement (appendix A) and complete a questionnaire to identify eventual parameters that could influence their results. The students then took a pre-test to evaluate their initial knowledge. The students were then invited to complete two learning activities to become familiar with the vocabulary; these learning activities were the same for both groups. Afterward, the students had a study session on the SRS attributed to their group. During the second lesson, the students trained again on the same SRS and then completed an
immediate post-test. Two weeks later, the students took a delayed post-test to evaluate their vocabulary acquisition.

3.3.1 Questionnaire
The students were invited to fill out a questionnaire (see appendix B) which asked about their age, mother-tongue, any additional languages spoken, and gender prior to the experiment to identify the students’ different particularities that could impact the study’s outcome.

3.3.2 Selection of the Target Words
For this study, 18 adjectives have been selected as the target words. They were selected from two vocabulary exercise books designed for students aiming to reach the C1 level, to ensure the relevance of these words for the learners. The following books have been used “Destination C1 & C2 Grammar and Vocabulary” by Mann and Taylore-Knowles (2010), as well as “The Vocabulary Files – English Usage Advanced” by Bestis & Mamas (2010). When selecting vocabulary items, special attention has been directed towards the students’ unfamiliarity with the target words. It was indeed necessary that the students were unlikely to know the target words to preserve the integrity of this study. When selecting the words, a pattern emerged, as many of those words were adjectives. Consequently, the decision was made only to include adjectives in the present study. This choice can be motivated by limiting the study to a category of words, which can increase the noticeability of the results (Folse, 2006).

Another parameter was the proximity of these adjectives to their equivalent in the Swedish language; their similarity to Swedish could have impacted the study’s outcome, so the eventual adjectives possessing proximity to the Swedish language were not selected. Another parameter was variety. Indeed, the chosen adjectives had different forms and spellings to show the diversity of descriptive words. The adjectives were also chosen to give the participants different degrees of difficulty. Indeed, the chosen adjectives cover a wide range of difficulty levels, with some adjectives that might be simpler and some complex ones to accommodate the heterogenic level of the participants.

3.3.3 Pre-test
Before treatment, a pre-test focusing on the 18 adjectives was administered to assess the students’ vocabulary knowledge. This test aimed at establishing a baseline of the students’ knowledge before the treatment, which was helpful to compare the learners’ performance before and after treatment and evaluate the impact of the latter on the students’ vocabulary acquisition; 25 minutes were allocated to that pre-test for the students not to feel pressured by time constraints. The test was identical for all the students, independently of their group, Socrative or Gimkit. This test was similar in structure to both post-tests. However, it differed slightly in some respects. Indeed, the order in which test items came in each test was randomized to limit the recall effect that could have occurred if the order in which the test items were presented remained the same.
from one test to another. However, the tests remained with the same structure to ensure that the difficulty level of each test remained comparable. Indeed, comparability is essential as more complex or accessible post-tests could generate artificial results (Mackey & Gass, 2021).

3.3.4 Study Sessions

During the second lesson, the students trained for 20 minutes on the SRS they had been assigned (Socrative or Gimkit) before taking the immediate post-test. The students were allowed to do the test on Socrative as many times as needed, and Gimkit has a functionality that enables the teacher to choose the desired training time, which was 20 minutes. When training on the different SRSs, the students trained on their receptive and productive competencies as they were asked to fill in gaps by choosing the right words for the sentence, as well as true of false questions testing the students on the meaning of the words, as well as attributing the correct definition to a word. All these exercises were aimed at preparing the students for the tests that would take place shortly after and, therefore, resemble the test items in their shape and vocabulary frequency. As Mondria & Wiersa (2004) explain while referring back to Griffin’s study (1992), the equivalence between the type of learning and the type of test is central to get better results, therefore highlighting the importance of having the tests correspond closely to the training the students receive.

3.3.5 Immediate Post-test and Delayed Post-test

As this study involved two classes, the tests were not conducted on the same day but were administered at equal intervals between sessions. The immediate post-test occurred during lesson two, and the delayed post-test occurred during lesson three, which took place two weeks later. Those times have been chosen for practical reasons, aiming not to impede the teacher’s progress within the current unit they are teaching to the students. As Mackey and Gass explain (2021), learning is a process and does not usually occur only once; therefore, a delayed post-test sometime after the immediate post-test can help evaluate the treatment's long-term effects.

The students took the immediate post-test shortly after training on the digital tool attributed to their group. Nation (2013) argues that successfully recalling the item enhances retention rather than being passively exposed to it (Nation, 2013). Therefore, using digital tools to study vocabulary items is a cohesive choice.

3.3.6 Data Analysis

The data from the three tests was statistically analyzed from a quantitative point of view. For this, the software package SPSS Statistics was used. This software is widely used for statistical analysis across diverse fields like social sciences, market research, and healthcare. It supports various statistical procedures, data management, and visualization tools.
SPSS can be used to check the assumptions of the chosen statistical analysis method, such as normality and homogeneity of variance. Moreover, it can be used for mathematics to interpret results, analyze the output provided, including the main and interaction effects, and interpret the significance of the findings.

3.3.7 Socrative

Socrative is an SRS that can be used both gamified and non-gamified. For this study, Socrative was used in a non-gamified mode. When using this SRS, the teacher can decide if the students get immediate feedback and whether the quiz is student- or teacher-paced. Another asset of this tool is that the teacher can actively monitor the students’ answers as they take the test; those results will remain visible to the teacher after the study session, enabling them to adapt the upcoming lessons if some notions need to be studied again. The teacher can use multiple-choice, true or false, or fill-in-the-blank questions in this SRS. The test items can be presented to the students in the same order as the teacher entered them, or they can be introduced randomly to the students, depending on the teacher’s choice. Moreover, the non-gamified version does not permit the students to see and compare their scores, and does not contain competitive features.

3.3.8 Gimkit

Gimkit was used in the classic mode. This SRS is gamified, as each good answer allows the students to earn coins, as visible in Figure 1 below. Those coins can then be used to buy power-ups and upgrades, as displayed in Figure 2 below. These power-ups can be used to multiply the number of coins earned during the next question, sabotage another player, or change the screen’s appearance. If the students answer several questions correctly in a row, they can receive more coins as a reward.

In this SRS, the student answers some multiple-choice and fill-in-the-blank questions. Therefore, the kind of questions allowed is very similar to those in Socrative, which is one of the criteria that motivated the choice of this SRS, in particular for the present study. The test items were presented randomly to the students, independently of the order in which they were entered. At the end of the test, the teacher can consult a report to identify which questions counted the highest number of incorrect answers and which had the highest number of correct answers, thereby enabling the teacher to adapt the upcoming lessons depending on the results. Gimkit displays the students’ points in real time, creating a competitive frame for the students’ learning.
3.4 Design of the Tests

The test items in this study were based on Nation’s *Vocabulary level tests* (Nation, 2013), as Nation’s work is well recognized in the field. The tests were designed to examine the participants’ knowledge and acknowledge even limited word knowledge (Nation, 2013). According to Nation (2013), a good vocabulary test evaluates and positively impacts learning. The test was aimed at testing the students’ knowledge of 18 adjectives studied previously. Each test was divided into four exercises, later called *test items*. Each test contained the same test items.

The first test item consists of filling in the gap sentence in which the students must use the newly acquired vocabulary to fill in the sentence. This exercise tests the participants productively and enables them to use the vocabulary in a practical context, which requires them to understand the meaning and usage of the words in the said context to use them appropriately. A few letters at the beginning of each word have been written to prevent the students from filling in the blank with words other than those tested (Laufer & Nation, 1999). An example of this test item is the following:

<table>
<thead>
<tr>
<th>Complete the underlined words.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The café had an abu________________________ variety of freshly brewed coffees, satisfying every caffeine craving.</td>
</tr>
</tbody>
</table>

The second test item included are true or false questions designed to gauge the students’ comprehension of the target words. Indeed, the students received five true or false statements based on the words’ definitions and meanings and had to identify whether the statement was true or false. This test item is named in Nation (2013) and is pointed out as commonly used in vocabulary tests. The test item was presented as follows:
Indicate whether the statement is true or false. If you do not understand the sentence, leave the answer blank.

Using the word equidistant to describe two points means that those points are at different distances from a reference point.

TRUE  FALSE

Figure 4. Example of a true or false statement extracted from the immediate post-test (Appendix E).

The third test item was a multiple-choice item with definitions based on the format found in the levels test of Nation (2013). Some distractors were included in this exercise. This test item assesses the students’ receptive knowledge of the words by making the students match the correct definition to the chosen words. Some distractors were included in this exercise. An example of this test item is the following:

Choose the word that matches the definition.
Inattentive – Prejudiced – Deliberate

A) _______________: showing an unreasonable dislike for something or someone.
B) _______________: not giving attention to someone or something.

Figure 5. Example of a multiple-choice test item extracted from the pre-test (Appendix C).

One issue encountered during the elaboration of the tests was ensuring the test comparability. Indeed, it was crucial to ensure that the tests’ difficulty levels did not significantly differ from one test to another as it could influence the study’s outcome. However, using the same test as a pre-test, immediate post-test, and delayed post-test would not have been appropriate research practice. This approach could impact the results, as the students might become familiar with the questions or show reduced attentiveness (Mackey & Gass, 2021). To mitigate that effect, the sentences between each test were modified so as not to seem too familiar to the students. However, special attention was allocated to the structure of the test so it would remain the same. To ensure the equivalence of the tests, the frequency of the different lexical items has been compared using the Corpus of contemporary American English, as advised by Mackey & Gass (2021). Another essential matter to consider was the directions of learning and testing. Learners tend to recall information more effectively when the learning method aligns with the testing method (Nation, 2013). Therefore, for this study, words practiced productively were tested productively, and words practiced receptively were tested receptively.

3.5 Ethical Considerations

For the present study, the students have been asked to fill out a consent form (Appendix A) in accordance with the university’s policy (2022), and an e-mail has been sent to the legal guardians to inform them of the study. The students were informed that their
participation in the study was anonymous and that their name or identity would not be displayed.

4. Results

A total of 61 high school students participated in the study. The participants were divided into two groups corresponding to their respective classes, and each class formed a distinct group. Each of these two groups was exposed to a different condition, meaning a type of vocabulary learning environment and training method. In one group, the participants were exposed to Socrative, which has no gamification elements. The other group was exposed to Gimkit, which contains gamification elements and possesses an environment fostering competitiveness.

The study took place over a period of three weeks and exposed the students to different conditions of vocabulary training (depending on their group) at two points in time; these events will later be referred to as study sessions. During this period, the dependent variable, test score, was controlled at three points in time: before the study in a pre-test, directly after the second vocabulary training in an immediate post-test (which will be referred to as mid), and at the end of the study two weeks after any training in a delayed post-test (which will be referred to as post). After completing all three tests, only 37 participants from both groups completed the three-week study in full; hence, only 19 participants remained in the control group and 18 participants in the gamification group.

The software program SPSS was used to analyze the results of each group’s three tests. Indeed, descriptive statistics are generally used to get an overview of the data. SPSS helped summarize and explore the results, providing valuable insights into the data’s central tendency, variability, and distribution. Analyzing descriptive statistics is the foundation for more advanced statistical analyses and hypothesis testing. Table 3 provides an insight into the descriptive statistics of the full sample.

Table 3. Descriptive statistics of the full sample divided into groups and tests.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socrative</td>
<td>Pre 19</td>
<td>5.3</td>
<td>2.8</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Mid 19</td>
<td>11.9</td>
<td>4.7</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Post 19</td>
<td>10.8</td>
<td>4.1</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>Gimkit</td>
<td>Pre 18</td>
<td>6.7</td>
<td>2.7</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Mid 18</td>
<td>12.7</td>
<td>5.9</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Post 18</td>
<td>11.7</td>
<td>5.4</td>
<td>2</td>
<td>18</td>
</tr>
</tbody>
</table>

From analyzing the descriptive statistics of the full sample comprising both groups (Table 3), an increase in the mean score can be noticed. Indeed, for the group “Socrative”, the mean went from 5.3 in the pre-test to 11.9 and 10.8 (out of the maximum score of 18) in the mid- and post-test, respectively. As for the group “Gimkit”, the mean went from 6.7 in the pre-test to 12.7 and 11.7 in the mid- and post-test, respectively. The lower mean
value for the pre-test suggests that, on average, participants scored fewer points in their initial test compared to the mid- and post-test. The increasing mean values for the upcoming two tests (mid and post) indicate a positive trend. On average, participants scored higher in the subsequent tests, suggesting a potential improvement in vocabulary skills when exposed to Socrative and Gimkit.

Upon analyzing the standard deviation within the sample, a lower standard deviation for the pre-test can be noticed (see Table 3). Specifically, for the group “Socrative”, the pre-test demonstrates a standard deviation of 2.8, whereas the mid-test and post-test show a deviation of 4.7 and 4.1, respectively. Similarly, for the group “Gimkit”, the pre-test shows a deviation of 2.7, while the mid-test and post-test display deviations of 5.9 and 5.4. The participants’ scores were more closely grouped around the mean in the initial tests for both groups, and the increase in standard deviation for the following two tests implies that individual scores became more spread out from the mean. This suggests a greater variability in vocabulary knowledge among participants in the subsequent tests, with some participants showing higher and others lower levels of improvement from the training.

To summarize, upon analyzing the descriptive statistics, the results suggest an overall improvement in vocabulary knowledge over the three tests. However, the increased standard deviation in the subsequent tests indicates greater variability in individual performance compared to the more consistent outcomes observed in the initial test. However, when compared, no clear distinction in test results between each group - Socrative and Gimkit - could be seen at any particular test. More advanced statistical analyses and hypothesis testing will follow to determine the significance of results and causality between groups.

To determine which statistical method to use, first, the normality of the data had to be checked. As Larson-Hall (2010) explains, parametric tests (e.g., t-tests and ANOVA) assume a normal distribution, while non-parametric tests (e.g., Friedman test and Kruskal-Wallí’s test) are less sensitive to distributional assumptions. The Shapiro-Wilk test was used to test for normality, which was especially suitable in this context of small samples, as Ricci (2005) recommends. The null hypothesis of the test was that the population was normally distributed; thus, if the p-value was less than the chosen significance level (0.05), the null hypothesis was rejected, suggesting that the data did not come from a normally distributed population. Table 4 below displays the results of the test of normality executed on the two groups.

Table 4. The results from the test of normality – the Shapiro-Wilk test.

<table>
<thead>
<tr>
<th>Test</th>
<th>Group</th>
<th>Shapiro-Wilk Statistic</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>Socrative</td>
<td>0.940</td>
<td>19</td>
<td>0.267</td>
</tr>
<tr>
<td></td>
<td>Gimkit</td>
<td>0.940</td>
<td>18</td>
<td>0.293</td>
</tr>
<tr>
<td>Mid</td>
<td>Socrative</td>
<td>0.948</td>
<td>19</td>
<td>0.361</td>
</tr>
<tr>
<td></td>
<td>Gimkit</td>
<td>0.798</td>
<td>18</td>
<td>0.001</td>
</tr>
<tr>
<td>Post</td>
<td>Socrative</td>
<td>0.890</td>
<td>19</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>Gimkit</td>
<td>0.908</td>
<td>18</td>
<td>0.079</td>
</tr>
</tbody>
</table>
When examining Table 4, showing the result from the Shapiro Wilk test, the null hypothesis is rejected for the mid-test in the “Gimkit” group and for the post-test in both groups. This rejection occurs because the \( p \)-value (Sig.) is less than 0.05. Even though some of the test results show significance, it does not clearly show that the data comes from a normally distributed population. Hence, due to all the data not being normally distributed, a non-parametric statistical method was used in accordance with Larson-Hall’s (2010) recommendation.

The Friedman test is the non-parametric alternative to repeated measures ANOVA, used when comparing three or more related groups (in this context, the three within-subjects factor of time). The dependent variable is ordinal or continuous but not normally distributed. If the Friedman test indicates statistical significance, it suggests differences between the groups. An important remark, however, is that the test does not identify which specific groups differ from each other. Table 5 below displays the results of the Friedman test.

<table>
<thead>
<tr>
<th>Table 5. The results from the Friedman test.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Chi-Square</td>
</tr>
<tr>
<td>df</td>
</tr>
<tr>
<td>Asymp. Sig.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Chi-Square</td>
</tr>
<tr>
<td>df</td>
</tr>
<tr>
<td>Asymp. Sig.</td>
</tr>
</tbody>
</table>

The results of the Friedman test, displayed in Table 5, indicate a statistical significance. The \( p \)-value is less than the significance level of 0.05, and the test’s null hypothesis is rejected. Hence, there is a difference in the distributions of vocabulary scores across the three tests. In other words, it can be statically shown that the scores of both groups were improved upon vocabulary training using the digital tools Socrative and Gimkit. However, as mentioned earlier, the Friedman test cannot confirm causality and the relationship between groups. If the Friedman test indicates statistical significance among groups, post hoc tests are needed to identify where the differences lie or if there are any differences between the two groups (Larson-Hall, 2010).

The Wilcoxon signed-rank test is a post hoc test comparing two related groups and determining whether there is statistical significance between them (Larson-Hall, 2010). The Wilcoxon Signed-Rank test is used to pinpoint which specific pairs of groups are significantly different. The three tests were assembled into pairs, comparing each combination. When analyzing the result, the significance level was adjusted using Bonferroni correction to control the error rate due to multiple comparisons. The adjusted significance level became \( 0.05 / 3 = 0.0167 \) for each comparison. Table 6 introduces the results of the Wilcoxon signed rank test:
Table 6. The results from the Wilcoxon signed-rank test.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mid - Pre</th>
<th>Post - Pre</th>
<th>Post - Mid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socrative</td>
<td>Z</td>
<td>-3.688(^a)</td>
<td>-3.736(^a)</td>
</tr>
<tr>
<td></td>
<td>Asymp. Sig. (2-tailed)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Gimkit</td>
<td>Z</td>
<td>-3.249(^a)</td>
<td>-3.165(^a)</td>
</tr>
<tr>
<td></td>
<td>Asymp. Sig. (2-tailed)</td>
<td>0.001</td>
<td>0.002</td>
</tr>
</tbody>
</table>

\(^a\) Based on negative ranks.  
\(^b\) Based on positive ranks.

When examining Table 6, showing the results from the Wilcoxon signed-rank test, if the \(p\)-value for a specific pair is less than the adjusted significance level (0.0167), the null hypothesis is rejected for that pair, therefore indicating a statistical significance in vocabulary scores between those two tests. Statistical significance exists in both groups for the pairs: pre-test and mid-test, and pre-test and post-test. However, the pair mid-test and post-test indicate that the \(p\)-values are greater than the adjusted significance level: \(p = 0.137\) for the Socrative group and \(p =0.193\) for the Gimkit group. Hence, for both groups, there is no statistical significance in vocabulary scores between the two tests (mid- and post-tests). Even though the mid-test and post-test comparison is not statistically significant, there might still be a meaningful change in scores that the statistical tests have not detected. It might instead indicate that the vocabulary was kept long-term between the time interval of the mid-and post-test.

Analyzing these results may suggest a significant improvement in scores from the pre-test to both the pairs mid-test and post-test, indicating learning effects in both groups using Socrative and Gimkit. However, there might not be a significant change in scores between the mid-test and post-test, suggesting a plateau or stabilization of long-term knowledge retention in both groups. Hence, both groups showed statistical significance.

For a more comprehensive understanding, the ranks table from the results of the Wilcoxon signed-rank test, visible in Table 7, provides additional insights into the distribution of differences between the groups and the pairs. This table helps understanding the statistical significance, direction, and magnitude of changes in vocabulary scores.
Table 7. The results from the Wilcoxon signed-rank test using a ranks table.

<table>
<thead>
<tr>
<th>Group</th>
<th>Ranks</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socrative</td>
<td>Mid - Pre</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative Ranks</td>
<td>1a</td>
<td>3.50</td>
<td>3.50</td>
</tr>
<tr>
<td></td>
<td>Positive Ranks</td>
<td>18b</td>
<td>10.36</td>
<td>186.50</td>
</tr>
<tr>
<td></td>
<td>Ties</td>
<td>0c</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post - Pre</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative Ranks</td>
<td>0d</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Positive Ranks</td>
<td>18e</td>
<td>9.50</td>
<td>171.00</td>
</tr>
<tr>
<td></td>
<td>Ties</td>
<td>1f</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post - Mid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative Ranks</td>
<td>10g</td>
<td>7.60</td>
<td>76.00</td>
</tr>
<tr>
<td></td>
<td>Positive Ranks</td>
<td>4h</td>
<td>7.25</td>
<td>29.00</td>
</tr>
<tr>
<td></td>
<td>Ties</td>
<td>5i</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gimkit</td>
<td>Mid - Pre</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative Ranks</td>
<td>3a</td>
<td>3.67</td>
<td>11.00</td>
</tr>
<tr>
<td></td>
<td>Positive Ranks</td>
<td>15b</td>
<td>10.67</td>
<td>160.00</td>
</tr>
<tr>
<td></td>
<td>Ties</td>
<td>0c</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post - Pre</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative Ranks</td>
<td>3d</td>
<td>4.33</td>
<td>13.00</td>
</tr>
<tr>
<td></td>
<td>Positive Ranks</td>
<td>15e</td>
<td>10.53</td>
<td>158.00</td>
</tr>
<tr>
<td></td>
<td>Ties</td>
<td>0f</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post - Mid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative Ranks</td>
<td>8g</td>
<td>8.00</td>
<td>64.00</td>
</tr>
<tr>
<td></td>
<td>Positive Ranks</td>
<td>5h</td>
<td>5.40</td>
<td>27.00</td>
</tr>
<tr>
<td></td>
<td>Ties</td>
<td>5i</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td>18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Mid < Pre  
b. Mid > Pre  
c. Mid = Pre  
d. Post < Pre  
e. Post > Pre  
f. Post = Pre  
g. Post < Mid  
h. Post > Mid  
i. Post = Mid

Table 7 above shows the results from the test using a ranks table. This table divides the paired tests and compares each individual’s results into categories showing a negative, positive, or tied result. The pair pre-test and mid-test ranks reveal that 18 out of 19 participants in the Socrative group and 15 out of 18 in the Gimkit group displayed an
improvement. Similar findings also emerged for the pre-test and post-test pairs, with 18 out of the participants in the Socrative group and 15 out of 18 in the Gimkit group, which showed a positive improvement. Notably, one participant in the Socrative group did not improve after the vocabulary training, and three participants in the Gimkit group experienced a negative impact after the vocabulary training. Regarding the pair mid-test and post-test, the majority got a negative rank in both groups, showing 10 and 8 individuals, respectively.

Relating the above result to the research question of the present study: “How does a gamified learning environment in SRS impact vocabulary acquisition?” and the corresponding hypotheses, the answer is varied. From a statistical point of view, there is a statistical significance of time - short-term and long-term retention - on vocabulary acquisition using SRSs. Hence, using both Socrative and Gimkit in vocabulary learning contributes positively to the students’ vocabulary acquisition. However, the results also showed no statistical significance between the two groups. Hence, a gamified learning environment using SRSs was shown to have no statistical significance on vocabulary acquisition compared to a non-gamified one. Moreover, due to the non-normal distribution of the data and the entailing non-parametric model, the hypothesis relating to the interaction effect between the setting and time cannot be analyzed statistically with certainty. Hence, it will instead have to be analyzed descriptively in the discussion.

5. Discussion

Learning vocabulary is a longitudinal process requiring time; as Murre and Dros (2015) explained, repetition is critical as newly acquired information tends to disappear shortly after the first learning session. Therefore, training on vocabulary using an SRS can be an efficient way to practice. Moreover, SRS can provide immediate feedback, which, according to Winne & Hadwin (1998), can help the learners become aware of the knowledge that needs to be studied again. These two elements could, therefore, explain the positive impact of SRS on learning that has been uncovered in the present study.

The results of this study reveal that both Socrative and Gimkit contributed positively to the students’ vocabulary acquisition. These results are in accordance with previous studies using SRSs in the educational context (Patterson et al., 2010; Zhou, 2016). The SRSs contributed to the acquisition of receptive and productive vocabulary, reflecting the findings of Behbahani & Shahbazi (2022), who demonstrated that games contribute positively to learning receptive and productive vocabulary.

The present study also pointed out that no statistical significance was found between the students who learned vocabulary with a gamified tool and the ones who studied with a non-gamified tool, which is a point that has yet to be brought up in studies. In contrast, as Rasti-Behbahani & Shahbazi (2022) and Vandercruysse et al. (2013) showed in their studies, gamified learning tends to positively impact the learners’ motivation, impacting their learning acquisition. However, the results of the present study did not suggest that a gamified and competitive setting results in better vocabulary acquisition. Salomon (1984) nuanced this perspective by mentioning the negative impact that gamification can have...
on learning. The results of the present study do not align with this notation, as both student groups who trained with gamification and without gamification obtained similar results. This indicates that gamification did not have a statistically significant negative impact on these students’ learning. Gamification can, however, potentially positively impact the students’ learning outcomes and motivation; hence, reminding the students of the expected learning outcomes of the activity can be beneficial to reinforce the activity’s significance. Due to the short time span of this study, which prevented the possibility of a longitudinal study, it was not feasible to statically analyze the interaction effect between the setting and time. However, what is hypothesized is that gamification might improve motivation during study sessions – which in turn might lead to improvement in long-term retention. This is supported by the study of Vandercruysse et al. (2013), which explained that students exercising in a gamified learning environment had more appetite to the task and displayed higher enjoyment.

Several limitations in this study need to be acknowledged; the first limitation that can be acknowledged was that this study solely englobed adjectives, limiting the generalizability of the results to the word categories such as nouns and verbs. This constraint should be taken into account when applying the findings of the present study to a broader context. Moreover, the relatively short time span of the study might have influenced the outcomes as a longitudinal study might have provided deeper insight into the impact of motivation on learning.

This study shed some light on the positive effects of both Gimkit and Socrative. However, further exploration of the topic is required. Indeed, further studies could include different word categories, have a longer time span, or even include students of different ages. This could improve teaching methods and help teachers optimize how SRSs are included in teaching.

6. Conclusion

This study aimed to analyze how a gamified learning environment in SRSs impacts vocabulary acquisition. In this study, two main effects and one interaction effect were analyzed and hypothesized. For the first main effect of time, a significant effect on vocabulary acquisition was observed, indicating a retention both in the short-term and long-term using SRSs. However, for the second main effect, the learning environment, the results showed no statistical significance between the studied groups: gamified and non-gamified. Thus, a gamified learning environment does not yield greater benefits for vocabulary acquisition compared to a non-gamified one. Due to the limitations of the non-normally distributed data, no statistical conclusion of the interaction effect could be drawn for the interaction effect between environment setting and time.

This study provided insight into how both SRSs contributed positively to the students’ vocabulary acquisition. While the study did not show statistical significance between the two groups, a more extended study evaluating the students’ vocabulary acquisition throughout the school year enrolling in more classes could have brought up more insights into the eventual differences between the two SRSs. Moreover, a similar study with
younger learners could have been interesting to see if a gamified approach would yield different results.

The implications of this study are multifaceted and provide significant insights into the use of gamified and non-gamified SRSs for vocabulary acquisition in the classroom. While the present study revealed a substantial positive effect of time on vocabulary retention, both in the short and long term, when using SRSs, it did not identify a significant difference between a gamified and non-gamified learning environment. This suggests that the gamification aspect itself in SRSs may not enhance vocabulary acquisition compared to a non-gamified approach. However, the present study did shed light into the positive impact of both gamified and non-gamified SRSs on the students’ vocabulary acquisition, emphasizing the effectiveness of these SRSs for vocabulary learning. This study therefore highlighted the fact that teaching does not always need to be entertaining to facilitate learning. However, as explained by Vandercruysse et al. (2013) previously, gamified learning can have a positive influence on the students’ motivation and impact their appetite for the task. Gamification can therefore be an asset for the classroom, and it is up to the teacher to get to know their students and choose the SRS that will fit them best.
References


Tan, P. M., & Saucerman, J. J. (2017). Enhancing learning and engagement through gamification of student response systems. 2017 ASEE Annual Conference & Exposition,


Appendix A

Informationsblad till deltagaren i studieprojekt om vokabulärinlärning i engelska, HT 2023

Mitt namn är Cindy Xerri, och jag är student inom lärarutbildning vid Stockholms universitet. För närvarande genomför jag en studie om vokabulärinlärning. I detta brev finner ni information and min studie och vad delta i den här studien innebär. Studien har godkänts av engelskaläraren Ingrid Heiming.

• Syfte och genomförandet av studien

Studien fokuserar på elevernas inlärning av vokabulär och hur den påverkas av digitala verktyg. Eleven som väljer att delta i studien kommer att få genomföra tre tester och kommer att närvara vid tre undervisningstillfällen. Inga förberedelser krävs för att delta i studien.

• Vad det skulle innebära för dig att delta i studien

Denna studie görs i samarbete med era engelskalärare XXXX och ingår därför som ett vanligt undervisningstillfälle. Deltagandet i studien är frivilligt. Du har rätt att inte delta eller att avsluta dina medverkan när som helst utan att det påverkar dina studier. Det är viktigt att notera att oavsett om ni väljer att delta eller inte kommer ni att delta i undervisningen som vanligt.

• Etik och sekretess:


Kontakinformation och ansvarig student för studien
Cindy Xerri
cixe3272@student.su.se

Studiens handledare
Andrew Cooper
andrew.cooper@english.su.se
Samtycke till att delta i studieprojekt om vokabulärinlärning i engelska, HT 2023

Jag har läst och förstått den information om studien som anges i dokumentet ”Informationsblad till deltagaren i studieprojekt om vokabulärinlärning i engelska, HT 2023”. Jag har fått möjlighet att ställa frågor och jag har fått dem besvarade.

☐ Jag samtycker till att delta i studien som beskrivs i dokumentet ”Informationsblad till deltagaren i studieprojekt om vokabulärinlärning i engelska, HT 2023”.

☐ Jag samtycker till att mina personuppgifter behandlas på det sätt som beskrivs i ”Informationsblad till deltagaren i studieprojekt om vokabulärinlärning i engelska, HT 2023”

<table>
<thead>
<tr>
<th>Plats och datum</th>
<th>Underskrift och namnfärdliggende</th>
</tr>
</thead>
<tbody>
<tr>
<td>…………………………………………………………………</td>
<td>…………………………………………………………………</td>
</tr>
</tbody>
</table>

29
Appendix B

Frågeformulär

Namn: ______________________________________

Klass: ______________________________________

Kön:
☐ Man
☐ Kvinna
☐ Icke-binär
☐ Annat

Modersmål:
☐ Svenska
☐ Annat ______________________________________

Behärskar du andra språk än svenska, engelska, och eventuellt annat modersmål än ovan?

Språk: ______________________________________

Språknivå:
☐ ☐ ☐ ☐
Grundläggande kunskaper Goda kunskaper Mycket goda kunskaper Flytande

Språk: ______________________________________

Språknivå:
☐ ☐ ☐ ☐
Grundläggande kunskaper Goda kunskaper Mycket goda kunskaper Flytande

Språk: ______________________________________

Språknivå:
☐ ☐ ☐ ☐
Grundläggande kunskaper Goda kunskaper Mycket goda kunskaper Flytande
Appendix C

Pre-test

Exercise 1: Complete the underlined words. Only use adjectives.

Example: John Mayors was a charismatic and good-looking mayor.

1. After switching to a new job, her career became increasingly pro________________________, and she has been making more money than ever.

2. His cau________________________ comments during the meeting left a few colleagues offended.

3. She believed that opening an umbrella indoors was ji________________________ and would bring her bad luck.

4. The café had an abu________________________ variety of freshly brewed coffees, satisfying every caffeine craving.

Exercise 2: Indicate whether the statement is true or false. If you do not understand the sentence, leave the answer blank.

<table>
<thead>
<tr>
<th>Statement</th>
<th>TRUE</th>
<th>FALSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. An affable individual is usually unfriendly and difficult to approach.</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>2. A diligent worker is someone who often procrastinates and is not committed to their tasks.</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>3. Feeling disillusioned means having a strong sense of belief and trust in something.</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>4. Making an inadvertent error implies that the mistake was made without any intention or awareness.</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>5. Being equidistant from two cities means that you are at the same distance from both cities.</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

The test continues on the next page.
Exercise 3: Choose the word that matches the definition.
Inattentive - Prejudiced - Deliberate - Derogatory - Imminent - Eminent

A) ________________: showing an unreasonable dislike for something or someone.
B) ________________: not giving attention to someone or something.
C) ________________: famous, respected, or important.
D) ________________: coming or likely to happen very soon.
E) ________________: showing strong disapproval and not showing respect.

Exercise 4: Complete the underlined words. Only use adjectives.

1. The child's **def**_________________________ behavior, such as refusing to share toys, can be challenging for parents.
2. Some people who cannot read or write are called **ill**_________________________ because they have trouble with words.
3. The TV program avoided showing anything **ind**_________________________, so it was suitable for all ages.
4. Please provide a **succ**_________________________ explanation of your favorite holiday destination in one or two sentences.
Appendix D

Activity 1

Looking at the sentences under, try to guess the meaning of the words.

1. The inadvertent mistake in the report led to some confusion among the team.
2. After the company’s financial troubles, many employees felt disillusioned about their future there.
3. The affable host greeted each guest with a warm smile and made them feel right at home.
4. The eminent scientist received numerous awards for their groundbreaking research in the field of medicine.
5. The prejudiced remarks made by the critic revealed their strong bias against the artist’s work, making it difficult for others to evaluate it fairly.
6. The diligent student always completed their homework well before the deadline.
7. Using derogatory language is not acceptable in a respectful conversation.
8. The inattentive driver caused a minor accident because they were distracted by their phone.
9. The storm clouds gathering in the sky signaled that an imminent thunderstorm was on the way.

Step 2: Try to guess the adjectives (written in bold) of the words from the above sentences:

1. Inadvertent could mean___________________________________________________
2. Disillusioned could mean__________________________________________________
3. Affable could mean_______________________________________________________
4. Eminent could mean______________________________________________________
5. Prejudiced could mean____________________________________________________
6. Diligent could mean_______________________________________________________
7. Derogatory could mean____________________________________________________
8. Inattentive could mean_____________________________________________________
9. Imminent could mean_____________________________________________________

Step 3: Take a look at the definition and match it with the proper adjective.

a) ____________ : Not Intentional.
b) ____________ : Coming or likely to happen very soon.
c) ____________ : Showing strong disapproval and not showing respect.
d) ______________: Famous, respected or important.

e) ______________: Not giving attention to someone or something.

f) ______________: Showing an unreasonable dislike for something or someone.

g) ______________: Careful and using a lot of effort.

h) ______________: Disappointed and unhappy because of discovering the truth about.

i) ______________: Friendly and easy to talk to.
Activity 2

Step 1: Looking at the sentences under, try to guess the meaning of the words.

1. The prosperous business owner expanded her company to new markets.
2. Her caustic remarks left a bitter taste in the room.
3. She felt like her favorite team was jinxed because they always lost when she watched their games.
4. The garden was filled with abundant, colorful flowers.
5. The speaker delivered a succinct presentation that captured everyone's attention.
6. His indecent behavior at the party shocked the other guests.
7. Despite his best efforts, he felt embarrassed and frustrated because he remained illiterate, unable to read or write simple sentences.
8. She gave a defiant glare to the teacher when asked to put away her phone in class.
9. The two cities were equidistant from the swimming pool.

Step 2: Try to guess the adjectives (written in bold) of the words from the above sentences:

1. Prosperous could mean__________________________________________________
2. Caustic could mean_____________________________________________________
3. Jinxed could mean______________________________________________________
4. Abundant could mean___________________________________________________
5. Succinct could mean_____________________________________________________
6. Indecent could mean____________________________________________________
7. Illiterate could mean___________________________________________________
8. Defiant could mean______________________________________________________
9. Equidistant could mean__________________________________________________

Step 3: Take a look at the definition and match it with the proper adjective.

a) ___________: More than enough.
b) ___________: Hurtful, critical, or intentionally unkind way of speaking.
c) ___________: Unable to read and write.
d) ___________: Morally offensive, especially in a sexual way.
e) ___________: Proudly refusing to obey authority.
f) ___________: Having or believed to bring bad luck.
g) ___________: Said in a clear and short way.
h) ___________: Places or points that have equal distances between them.
i) ___________: Successful, usually by earning a lot of money.
Appendix E

Immediate Posttest

Exercise 1: Complete the underlined words. Only use adjectives.

Example: John Mayors was a charismatic and good-looking mayor.

1. Please provide a succ ______________ explanation of your favorite hobby in one or two sentences.

2. Some adults go back to school to learn to read and write if they are ill ____________.

3. The child's def ______________ attitude, like refusing to eat vegetables, can be challenging for parents.

4. The movie was rated as appropriate for all audiences, with no ind ____________ content.

Exercise 2: Indicate whether the statement is true or false. If you do not understand the sentence, leave the answer blank.

<table>
<thead>
<tr>
<th>TRUE</th>
<th>FALSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Using the word equidistant to describe two points means that those points are at different distances from a reference point.

An inadvertent mistake is one that was made intentionally and with careful thought.

Someone who is diligent in their work is known for being hardworking and thorough.

If someone is disillusioned, it means they have a strong sense of belief and trust in something.

An affable person is generally friendly and easy to talk to.

The test continues on the next page.
Exercise 3: Choose the word that matches the explanation.

Inattentive - Prejudiced - Eminent – Derogatory - Imminent – Diligent

A) _________________: coming or likely to happen very soon.
B) _________________: showing strong disapproval and not showing respect.
C) _________________: showing an unreasonable dislike for something or someone.
D) _________________: not giving attention to someone or something.
E) _________________: famous, respected, or important.

Exercise 4: Complete the underlined words. Only use adjectives.

Example: John Mayors was a charismatic and good-looking mayor.

1. The bakery had an abu________________________ supply of freshly baked bread, and the smell was delightful.
2. His cau________________________ remarks about the movie upset some of the viewers.
3. Ever since they moved to the new city, their business has been pro________________________, and they’re making more money than ever.
4. She believed that owning a black cat was ji________________________ and would bring her bad luck.
Appendix F

Delayed post-test

Exercise 1: Complete the underlined words. Only use adjectives.

Example: John Mayors was a charismatic and good-looking mayor.

1. The farmer’s field had an abu___________________________ harvest of ripe, juicy tomatoes this year.
2. His cau___________________________ comments sometimes hurt people’s feelings because he doesn’t think before he speaks.
3. The small business grew into a pro___________________________ company, making more money and expanding its business.
4. He felt like he was always jin___________________________ because every time he made plans, something went wrong.

Exercise 2: Indicate whether the statement is true or false. If you do not understand the sentence, leave the answer blank.

6. An affable person is generally friendly and easy to approach. ☐ ☐
7. Someone who is disillusioned has a strong sense of trust and belief in something. ☐ ☐
8. Someone who is diligent in their work is known for being hardworking and committed to their task. ☐ ☐
9. Using the word equidistant to describe two points means that those points are at the same distance from a reference point. ☐ ☐
10. An inadvertent mistake is one that was made unintentionally and without careful thought. ☐ ☐

The test continues on the next page.
Exercise 3: Choose the word that matches the explanation.

Derogatory - Imminent - Inattentive - Prejudiced - Eminent – Diligent

A) ___________________: showing an unreasonable dislike for something or someone.

B) ___________________: showing strong disapproval and not showing respect.

C) ___________________: not giving attention to someone or something.

D) ___________________: famous, respected, or important.

E) ___________________: coming or likely to happen very soon.

Exercise 4: Complete the underlined words. Only use adjectives.

Example: John Mayors was a charismatic and good-looking mayor.

1. The child’s def________________________ behavior, like not wanting to wear a hat on a cold day, can be challenging for parents.

2. People who cannot read or write are often called ill________________________.

3. Can you give me a succ________________________ explanation of your favorite color?

4. This TV show is suitable for all ages; there is nothing ind________________________ in it.