Reducing Forgetting Rate in EFL Students Using a Spaced Repetition-Powered Digital Game-Based Learning Application

Farah Aulia Rahman, Talitha Disti Amalia, and Muhammad Lutfi

Yayasan Teknologi untuk Indonesia, Indonesia

Author Note

All authors are currently at the research and development department, Yayasan Teknologi untuk Indonesia, Bandung, West Java, Indonesia.

The authors would like to thank Melvin Zhang Zhiyong, PhD for his valuable assistance in the adoption of the empirical forgetting rate calculation and helpful reviews throughout the data processing stage.

The research was fully funded by Solve Education Foundation, Singapore.

Correspondence should be addressed to Farah Aulia Rahman., Yayasan Teknologi Untuk Indonesia, Graha Padasuka Blok E no. 7. E-mail: farah.a@solveeducation.org

Abstract

Spaced repetition has been widely investigated as an effective technique to enhance human learning. However, a limitation of previous research is that most previous studies on spaced repetition in the English as a foreign language (EFL) learning contexts were dominated by in-class implementation and relied on small-scale controlled experiments. To fill this gap in the literature, in the present study, we conducted a large-scale natural experiment on spaced repetition using data from a digital game-based learning (DGBL) application, Dawn of Civilization. Specifically, we examined learners' natural learning repetition patterns in the game app and their effects on the forgetting rate of the learned words. The data were analyzed based on the identified repetition patterns, and differences in the forgetting rate of these patterns were analyzed. The results revealed that learners who closely followed the recommended spaced repetition schedule of the game had significantly lower forgetting rates, indicating better knowledge retention.

Keywords: digital game-based learning (DGBL), English as a foreign language (EFL), spaced repetition, forgetting rate, knowledge retention

Reducing Forgetting Rate in EFL Students Using a Spaced Repetition-Powered Digital Game-Based Learning Application

Digital game-based learning (DGBL) provides learners with an immersive learning environment where they can learn and/or test their knowledge using gaming through personal computers (PC), tablets, or smartphones. In recent years, due to its high level of mobility and flexibility, DGBL has become increasingly popular, including among English as a foreign language (EFL) teachers and learners (Hung et al., 2018; Lai & Zheng, 2018). Since many EFL learners are not immersed in the English-speaking environment, limited interactions in the EFL classroom can impede effective vocabulary acquisition and syntactic development. Addressing this need, DGBL offers ample opportunity for EFL learners to have engaging learning experiences, expand their English lexicon, and develop English language proficiency (Dehghanzadeh et al., 2019; Zhang et al., 2021).

Overall, it is common knowledge that, in order to become proficient in a foreign language, learners must acquire and retain a large amount of lexical items in their memory (Malone, 2018) while sustaining an active engagement with the learning tasks for a long period of time (Ortega, 2009). Obviously, such proficiency will not be achieved if most of the learned materials are forgotten (Seibert Hanson & Brown, 2020). To this end, DGBL offers the use of spaced repetition in the game. Spaced repetition is a technique where to-be-learned materials are distributed and repeated throughout the learning process using temporal intervals and interleaving techniques (Lee et al., 2021). In DGBL, learners no longer need to rely on a stack of cards to memorize the words or to manually schedule the repetition. In this sense, spaced repetition is the opposite of previously used massed repetition, or cramming technique, which is

defined as “a repeated practice without any temporal intervals between sessions and trials” (Suzuki & Hanzawa, 2021, p.1). To date, extensive evidence from cognitive psychology has been accumulated, showing that spaced repetition is the most effective technique to retain the knowledge for longer periods of time (Cepeda et al., 2006; Tabibian et al., 2019; Toppino & Gerbier, 2014).

However, a limitation of previous research on computer- or mobile-assisted spaced repetition in EFL learning is that most relevant studies have been dominated by smaller-scale in-class implementation or laboratory-controlled experiments (e.g., Chaikovska & Zbaravska, 2020; Seibert Hanson & Brown, 2020; Suzuki & Hanzawa, 2021; Zhang, et al., 2021). Accordingly, in the present study, following Tabibian et al. (2019), we investigated learners’ learning repetition patterns in a DGBL application, Dawn of Civilization, and their effects on forgetting rate and knowledge retention using a large-scale natural experiment. Dawn of Civilization is a non-commercial, multi-subject educational game application. One of the subjects taught through this application is the English language. Based on available evidence that integrating the spaced repetition approach in DGBL has a huge potential for optimizing EFL learners’ performance (Schimanke et al., 2017), in the present study, we aimed to address the following research questions:

1) What are the natural repetition patterns performed by players of Dawn of Civilization?
2) What are the effects of these natural repetition patterns on players’ forgetting rate?
3) Which repetition strategy is the most effective for improving players’ knowledge retention?

The remainder of this paper is structured as follows. Upon a literature review of previous studies on spaced repetition in EFL learning and the spaced repetition approach in DGBL, as well as a brief introduction to Dawn of Civilization, we present the method used in the present study. This is followed by the presentation of the results and the discussion section. The paper concludes with a summary of the main findings and an outline of directions of further research.

Literature Review

Spaced Repetition in EFL Learning

The concept of spaced repetition in learning can be traced back to the 19th century when Ebbinghaus (1885, as cited in Chukharev-Hudilainen & Klepikova, 2016) hypothesized that "the rate at which people forget information increased exponentially with time, but if an item was repeatedly revised, it tended to be forgotten at a slower rate, the latter gradually decreasing with each repetition" (p. 335). Since then, many studies have documented the advantage of implementing spaced repetition in learning including in second language (L2) and EFL contexts (e.g., Bloom & Shuell, 1981; Cepeda et al., 2006; Goossens et al., 2012).

All this body of past work confirmed that repeating learning materials using reasonable temporal intervals and interleaving techniques considerably assists learners' development of foreign language knowledge and skills (Lee et al., 2021; Seibert Hanson & Brown, 2020; Suzuki & Hanzawa, 2021). For instance, vocabulary learning using flashcard-based spaced repetition programs in the paired associate format was reported to increase retention of vocabulary knowledge relative to massed repetition (Nakata, 2015; Nakata & Suzuki, 2019; Zhu et al., 2012). Spaced repetition practice was also identified to be more effective than massed practice for the acquisition of receptive grammatical knowledge (Miles, 2014).

According to Toppino and Bloom (2002), learning is enhanced when a learner makes the most effort to repeat a previously learned material in a repeated practice session. The authors also argued that the repetition schedule should not be too short, massed, or even too long. In massed and short repetition learning, little effort is assigned to retrieving the material after the first learning session. Accordingly, all reviews or repetitions of the materials are crammed into one learning session (massed) or within an extremely short interval of time (short-spaced practice). However, if the repeat interval is too long, the material will very likely be forgotten.

Furthermore, other relevant studies suggested that the effectiveness of spaced repetition can be further optimized when the initially short interval between sessions gets gradually increased. This approach is also known as the lag effect (Melton, 1970; Roediger & Butler, 2011; Rogers & Cheung, 2020). However, previous findings on the lag effect remain largely inconsistent. For instance, while several studies empirically demonstrated the benefit of longer-spaced practice (Bird, 2010; Rogers, 2015) , others failed to replicate the difference in the effectiveness between shorter-spaced and longer-spaced practices (Kasprowicz et al., 2019), or even found that the shorter-spaced practice was more effective than the longer-spaced one (Suzuki, 2017; Suzuki & DeKeyser, 2017).

In essence, spaced repetition aims to enhance human learning (Tabibian et al., 2019). In the EFL learning contexts, for effective communication, learners should retain what they have learned for longer periods of time (Sanatullova-Allison, 2014). When a learner's fluency in a language reaches the point where conscious effort for language production (both verbal and written) is no longer needed, this means that the forgetting rate has been reduced, and knowledge retention has been achieved.

Spaced Repetition-Powered DGBL

In recent years, DGBL has revolutionized the landscape of EFL learning, particularly in the domain of vocabulary acquisition (Hung et al., 2018). Although, in learning English, it is extremely important to develop all comprehension and productive skills (i.e., reading, listening, writing, and speaking), learning English vocabulary can be particularly challenging due to the high demand of memorization. Accordingly, along with promoting learners' independent learning, collaboration, problem-solving (Kapp, 2012), and motivation (Fithriani, 2021; Waluyo & Bucol, 2021; Weissheimer et al., 2019), DGBL has been studied in terms of its capacity to make vocabulary learning more engaging (Kingsley & Grabner‐Hagen, 2018).

In this body of work, several algorithms to estimate the ideal lag time based on the learner's prior experience with the to-be-learned materials have been proposed. Three most remarkable of these algorithms are the Pimsleur Method (1967), Leitner's Learning Box (Leitner, 1972) and SuperMemo (Wozniak, 1990). These algorithms were discussed in various papers (e.g., Chukharev-Hudilainen & Klepikova, 2016; Seibert Hanson & Brown, 2020; Settles & Meeder, 2016; Tabibian et al., 2019; Zhang et al., 2021) and became the basis of many DGBL applications (e.g., Seibert Hanson & Brown, 2020; Settles & Meeder, 2016). As spacing intervals in these algorithms can be increased or decreased depending on student performance, they are considered to be adaptive.

Leitner's Learning Box, originally proposed by Leitner (1972) for the use with flashcards, is a spaced repetition algorithm where learning sessions are spaced with gradually longer intervals between them (Figure 1). The schedule for the flashcards in each box is different. Box 1 is reviewed every day; Box 2 is reviewed every other day; Box 3 is reviewed

every 4 days, and so on. The learning begins in Box 1, and when the flashcards are reviewed correctly, they are placed in the next box called "the promotion of intervals" (Zhang et al., 2021, p.5). Conversely, when the flashcard is reviewed incorrectly, it is placed back in the previous box regardless of which box the flashcard is currently in. Under this system, the length of time between reviews depends on how well a learner can remember the learning material: the better an item is remembered, the longer the time interval.

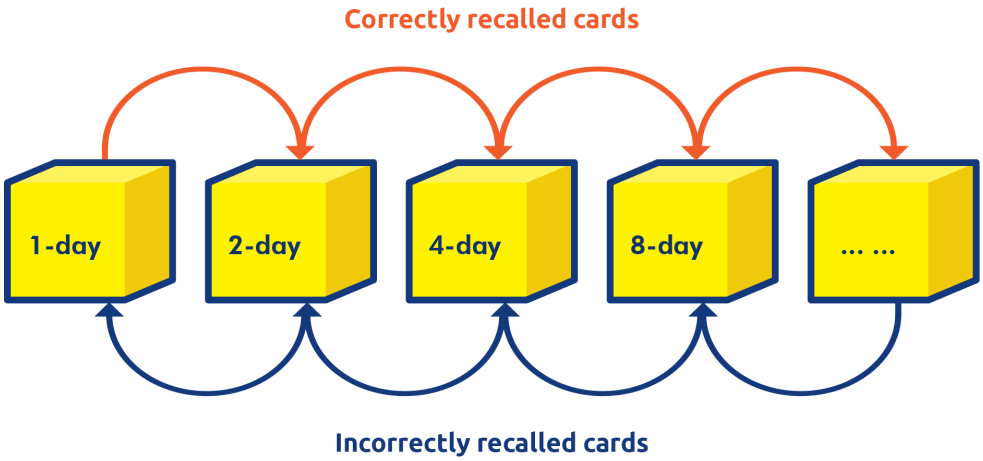


Figure 1. Leitner's box system

As in Leitner's Learning Box, spaced repetition-powered DGBL frequently involves the use of flashcards, multiple-choice questions, and spelling tests (see Godwin-Jones, 2011; Seibert Hanson & Brown, 2020; Tu et al., 2020; Zhonggen, 2018). Another common approach in spaced repetition-powered DGBL is to use pairing tests to help learners to associate words with their meanings (Nushi & Eqbali, 2017). These approaches are frequently applied separately or combined in a gaming environment. Furthermore, to improve learning engagement, spaced repetition-powered DGBL applies game elements such as levels, rules, competitions, and rewards (Tu et al., 2020; Zhang et al., 2021). In this context, levels refer to providing learners, upon their mastery of lower-level knowledge or skills, with tasks of increasing difficulty.

Furthermore, rules refer to having a clear presentation and explicit explanation of learning goals and ways. Next, competitions refer to providing learners with opportunities to interact and compete with peers to motivate and engage learners. Finally, rewards refer to rewarding students upon their completion of learning tasks for their successful performance (Prensky, 2003).

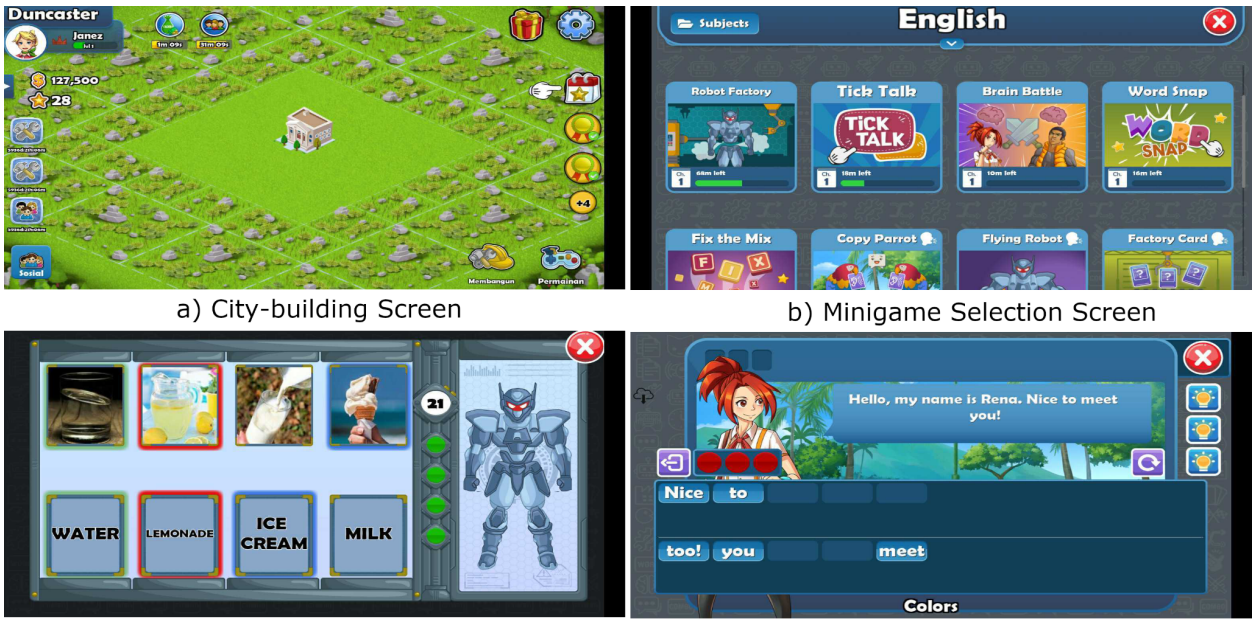
Dawn of Civilization

Dawn of Civilization[1] is a non-commercial DGBL application developed by Solve Education![2], a nonprofit organization that specializes in developing learning platforms aimed at particularly underserved communities. The game is available for mobile and desktop platforms and teaches various learning subjects, including EFL, via an immersive gaming environment (Kickmeier-Rust et al., 2007). In this environment, players are pulled into the world created inside the game to play, learn, and interact with other players. Dawn of Civilization allows players/learners to play the game either offline or online, free of charge—no subscription or in-app purchase is required.

Dawn of Civilization consists of the following two major parts: 1) a city-building game and 2) several mini-games that contain the lessons to support players' city-building journey. The city-building game is designed to be the fun part of the game to retain players so that they stay and continue learning. In this game, each player can create and manage a virtual city as a mayor. Players can construct buildings, place landmarks in their town, and earn taxes from their citizens. To develop the city, players will need resources and reward cards, which can be obtained by playing mini-games. It is through these quiz-based mini-games that EFL lessons are delivered. Available languages of instruction in the application are Indonesian, Hindi, and Thai.

[1] https://dawnofcivilization.net/
[2] https://solveeducation.org/

Figure 2(a) shows an example of an Indonesian-speaking player using an Indonesian language of instruction to learn EFL. When the player logs in to the game, s/he sees a city-building screen on which s/he can build his/her virtual city. As mentioned above, in order to collect resources for building, s/he must play the mini-games available in the game. Figure 2(b) shows the mini-game selection screen that lists 16 mini-games delivering EFL learning content for vocabulary building in isolation or through grammatically related words or concepts. The player taps a mini-game icon to access lessons of new material or to practice previously-learned material. Figure 2(c) shows a mini-game called "Robot Factory." In this mini-game, a player learns how to associate an image with a word or phrase. Figure 2(d) shows a mini-game called "Tick Talk" where a player learns how to form a logical sentence by putting the words in the right order.



a) City-building Screen

b) Minigame Selection Screen

c) Robot Factory Minigame

d) Tick Talk Minigame

Figure 2. Screenshots from Dawn of Civilization for an Indonesian-speaking user learning English on an Android phone in 2021: (a) a city-building screen: an area in the game for the player to build a virtual city on; (b) a mini-game selection screen: 16 mini-games with mini-lessons; (c) the "Robot Factory" mini-game that teaches new vocabulary in isolation through flashcards; (d) the screen of the mini-game called "Tick Talk" that teaches how to put the words in the right order to form a logical sentence.

Dawn of Civilization's players are allowed to tap and play any mini-game available on the mini-game selection screen. After completing a series of questions in one level in a mini-game, the player "unlocks" a new level that contains more advanced learning materials. Dawn of Civilization has the following six in-game levels: Pre-A, A1, A2, B1, B2, and C1 (see Figure 3). Depending on the results of the diagnostic mini-game that is placed before a learner accesses the mini-game selection screen, s/he can start from the easiest level (Pre-A) or any another level. The aim of this diagnostic mini-game is to incorporate into the game certain mastery learning elements (Block, 1971): before moving on to a more advanced learning level, any player must reach a certain level of prerequisite knowledge.

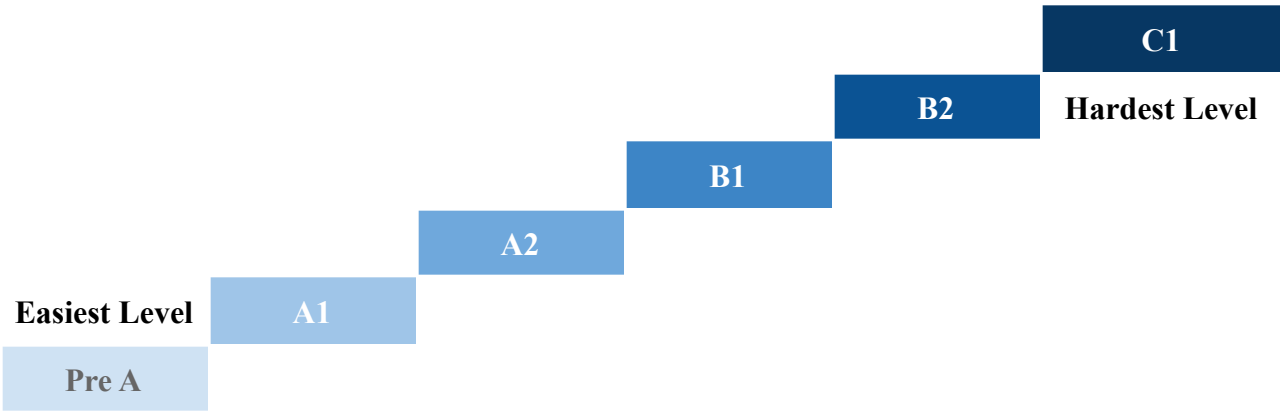


Figure 3. In-game leveling system for Dawn of Civilization (the EFL subject)

The content for Dawn of Civilization is created on Content+, a separate web-based application for teachers to create and deliver gamified learning materials through the application. Content creators can select which game template they want their materials to be delivered in through Content+. This system allows teachers to enter custom questions based on their curriculum, as well as to create a new subject comprising various mini-games and assign learners with the already existing subject in the game. The EFL subject in Dawn of Civilization was developed by Solve Education! (Amalia, 2018). To date, 16 mini-games containing vocabulary building, sentence structure, and some elements of listening and speaking are available; more mini-games are currently under development.

Teachers and content creators can input their content through Content+, a web-based content management system that allows learning content to be delivered through Dawn of Civilization (see Figure 4). When content is ready to publish in the game, the content package in Content+ is exported to a multi-purpose Customer Relationship Management (CRM) system where the developers perform sorting on all the available content to determine which region and on which platform (Android, iOS, personal computers, etc.) the content will be available in. The sorted content is then delivered to the Dawn of Civilization app through an Application Programming Interface (API) that acts as an intermediary to communicate which content needs to be collected from the CRM and then delivered to Dawn of Civilization.

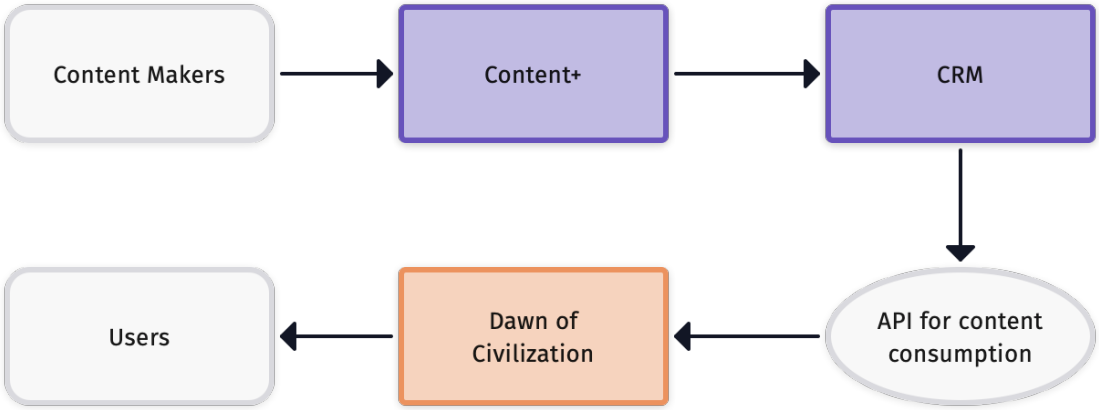


Figure 4. Creating learning content on Content+ and delivering it to Dawn of Civilization

Learning content on Dawn of Civilization is delivered via Leitner's Learning Box-inspired spaced repetition model called the Question Recycler (QR) system. Instead of being placed in boxes, to-be-learned materials (questions) in the game are assigned different repetition levels. Specifically, repetition level 1 (or QR level 1) is Box 1, QR level 2 is Box 2, and so on until QR level 5. At its first presentation to a player, any question is assigned an initial repetition level of 2 if answered correctly, or 1 otherwise. Furthermore, similarly to Leitner's Learning Box, in each subsequent attempt, repetition level will increase by 1 every time the question is reviewed correctly. Alternatively, the question is not returned back to repetition level of 1 (equivalent to returning the flashcard to Box 1 in Leitner's [1972] model), but its repetition level will only be decreased by 1 (equivalent to moving the flashcard to the preceding box in Leitner's [1972] model), unless the repetition level is already at 1. In this case, repetition level remains 1.

When the player attempts to recall a question at repetition QR level 4 and answers it correctly, the question earns QR level 5, and the user is considered to have completed the review cycle for this particular question. In that case, the question is taken out of the review cycle. The questions that have been taken out of the review cycle are no longer presented to the user unless there are no more questions available in the review cycle ready for review. Exceptions are made for questions on Pre-A level. Since questions on this level contain very basic learning materials (e.g., the alphabet, single-digit numbers, and simple words), fewer repetitions are assigned to Pre-A level. Accordingly, questions on this level are taken out of the review cycle when they reach repetition level 3 (instead of 5). Repetition level assigned to each learning item determines the recommended timing of the next review for the learner (see Table 1 below).

Table 1

Review Timing of Learning Items in the Game According to the QR System

	Review time	
	Pre-A Level	A1, A2, B1, B2, and C1 Levels
Repetition level = 1	Review in 1 day (24 hours)	Review in 1 day (24 hours)
Repetition level = 2	Review in 2 days (48 hours)	Review in 2 days (48 hours)
Repetition level = 3	N/A	Review in 4 days (96 hours)
Repetition level = 4	N/A	Review in 8 days (192 hours)

The repetition system in DOC has been programmed to present questions again to the player only when the waiting period is over.[3]

[33] Previous versions of the game allowed for an earlier presentation of an item to the player.

Method

The present study unfolded in the following three steps: 1) sampling; 2) identification of natural learning repetition patterns by Dawn of Civilization's users; and 3) data analysis.

Sampling

The dataset consisted of 355,190 study sessions, grouped into 115,651 unique player-question pairs collected from Dawn of Civilization over a one-year period (August 2020 - July 2021). In a single study session, depending on which mini-game players accessed the content through, they could be presented with between 1 to 6 questions. A question in Dawn of Civilization is the smallest unit of learning content that a player could consume. In the present study, all attempts of a single player to answer a specific question were grouped together and considered as a single data unit (thereafter referred to as player-question pairs). Only those player-questions pairs where the player made three or more attempts to answer a question were included in the dataset. Furthermore, since, in contrast to questions on other levels, questions on Pre-A level were required to reach a lower target QR level (i.e., fewer repetitions required to reach completion), the data for the second analysis—comparison of forgetting rates—were divided into the following two groups: (1) Pre-A (11,467 datasets) and (2) other levels combined (104,184 datasets). No differentiation in the data was made for the pattern identification analysis.

Pattern Identification

Based on the literature (Leitner, 1972; Suzuki & Hanzawa, 2021), we categorized the player-question pairs into the following four possible repetition behaviors:

1) massed repetition in which the players repeated materials in one learning session;

2) short-spaced repetition in which the players repeated materials earlier than the scheduled review time;
3) Leitner's (1972) repetition schedule in which the players repeated materials according to the scheduled time based on Leitner's Learning Box model;
4) long-spaced repetition in which the players repeated materials after the scheduled review time was over.

Dawn of Civilization's players learn organically—that is, they have complete control over their own learning schedule within the game. Accordingly, in order to understand the natural learning repetition patterns used by the players, each player-question pair was categorized into one of the four possible repetition patterns listed above. The definitions of each category that we relied on in the analysis were as follows:

1) massed repetition was defined as an attempt to answer the question made within 1 hour from the preceding attempt;
2) short-spaced repetition was defined as an attempt to answer the question made after 1 hour from the preceding attempt, but before 75% of the waiting period was over;
3) Leitner's (1972) repetition model was defined as an attempt to answer the question made within a range of ± 25% of the waiting period of the scheduled review time;
4) long-spaced repetition was defined as an attempt to answer the question made after 25% of the waiting period passed since the scheduled review time.

Upon categorizing all attempts into the four categories outlined above, we analyzed the categories of attempts for each player-question pair. A player-question pair was grouped into one of the repetition categories when most attempts for that player-question pair fell into the same

category. When none of the four categories stood out or if there was a tie between the number of attempts in more than one category, that player-question pair was excluded from the data. Overall, 40,344 player-question pairs could not be grouped into any of the four categories, which resulted in the inclusion of the remaining 75,307 player-question pairs for the subsequent pattern identification analysis.

Data Analysis

The data were analyzed to understand whether the four repetition categories had any impact on players' forgetting rates as an indicator of knowledge retention. Our specific focus was on identifying the repetition pattern with a lower forgetting rate.

To measure how likely the players would forget learning material they had just learned in the game, we used the empirical formula of forgetting rate previously proposed by Tabibian et al. (2019; see Eq. (1)). Here, the empirical forgetting rate is a measure of how fast the memory of an item decays at a certain point in time.

$$\hat{n} = -\log\left(\hat{m}\left(t_n\right)\right) / \left(t_n - t_{n-1}\right) \tag{1}$$

where $\hat{n}$ = forgetting rate; $\hat{m}(t_n)$ = value of answer (0.99 if correct, 0.01 if incorrect) at attempt n; $t_n - t_{n-1}$ = time interval between attempt n and the preceding attempt.

In the present study, attempt n was defined as the last attempt of the player-question pair where all preceding attempts had a recommended review time according to the spaced repetition system implemented in Dawn of Civilization. In the game, questions are only assigned a schedule for review when the target QR level required for completion has not been reached (see

the "Dawn of Civilization" section in Literature Review). Attempt *n* in which the empirical forgetting rate was calculated was an attempt after the target level was achieved as compared to its preceding attempt. It was at attempt *n* of each player-question pair that the empirical forgetting rate was calculated; this was done by taking the negative of the logarithmic value of the player-question pair at attempt *n* over the training interval (t_n - t_{n-1}).

The value for $\hat{m}(t_n)$ when a player had a correct recall was 0.99, while incorrect recall gave the value of 0.01. While, as mentioned above, the empirical forgetting rate formula used in the present study was adapted from Tabibian et al. (2019), our definition of a single data unit was different. Specifically, Tabibian et al. (2019) defined a single data as a set of questions testing the same word item; therefore, depending on the percentage of the correctly answered questions, the value of $\hat{m}(t_n)$ could range from 0 to 1. In contrast, a single data unit in Dawn of Civilization was assumed to consist of a single question, and the value of $\hat{m}(t_n)$ was binary (only correct = 1 or incorrect = 0). As the log of 1 is 0 and the log of 0 is undefined, all empirical forgetting rates (if calculated with the binary values) would be either 0 or undefined. Therefore, in order to avoid getting invalid values due to logarithmic function domain restriction—specifically, when the value was 0—the range for value for $\hat{m}(t_n)$ was trimmed on both ends by 1% from the initial range of 0 to 1. In the present study, forgetting rate was used as a dependent variable to evaluate which repetition patterns yielded better performance in terms of lower forgetting rate.

For a more fair comparison across items with potentially different levels of difficulty, the empirical forgetting rate was normalized using the median empirical initial forgetting rate of each item (see Eq. (2)).

$$\hat{n}_{0,\,(u,i)} = -\,log(\hat{m}\,(t_{(u,i),2})/\,(t_{(u,i),2} - t_{(u,i),1})\ (2)$$

where $\hat{n}_{0,\,(u,i)}$ = initial forgetting rate, and $t_{(u,i),2} - t_{(u,i),1}$ = time interval between second attempt and the preceding attempt for player-question pair

Using the same formula as that of the empirical forgetting rate (see Eq. (1)), the empirical initial forgetting rate was used to measure memory decay for a learning item after a single exposure to it ($\hat{n}_0$ was calculated at the second attempt of each player-question pair). Accordingly, relatively easy items would have lower empirical initial forgetting rates, whereas more difficult items would have higher values.

The median empirical initial forgetting rate of each item was then used to normalize the empirical forgetting rates of player-question pairs with corresponding items. In order to ensure that the empirical forgetting rate was not measured at the same attempt as the initial forgetting rate, only player-question pairs consisting of at least three attempts were included in the analysis.

The player-question pairs on Pre-A level were analyzed separately from other levels, because, as explained previously, spaced repetition of Pre-A questions stops on QR level 3 of Dawn of Civilizations, while questions on other levels stop at QR level 5. Accordingly, the same number of attempts on Pre-A level might not necessarily represent the same learning journey as compared to the other levels. For instance, four attempts on Pre-A level mean that a player made at least two mistakes before the question was considered complete, whereas four attempts on other levels mean that the player consecutively provided four correct answers. Finally, in order to identify the differences between median for normalized empirical forgetting rate among the groups, we initially planned to perform hypothesis testing using ANOVA. However, due to

unprecedented findings, we decided to group player-question pairs into two groups instead of four for the analysis of forgetting rates (see Results section for further details) and utilized the Mann-Whitney U test instead.

Results

Natural Repetition Patterns Used by Dawn of Civilization Players

As discussed in Literature Review, Dawn of Civilization uses hand-tuned spaced repetition algorithms inspired by Leitner's Learning Box that presents questions to players when the review time has been reached or is over. In the present study, since the players had full authority over their return to the game, they did not always review the learning materials presented in the game at specific recommended times. The results of natural learning repetition patterns observed in the sample are shown in Figure 5.

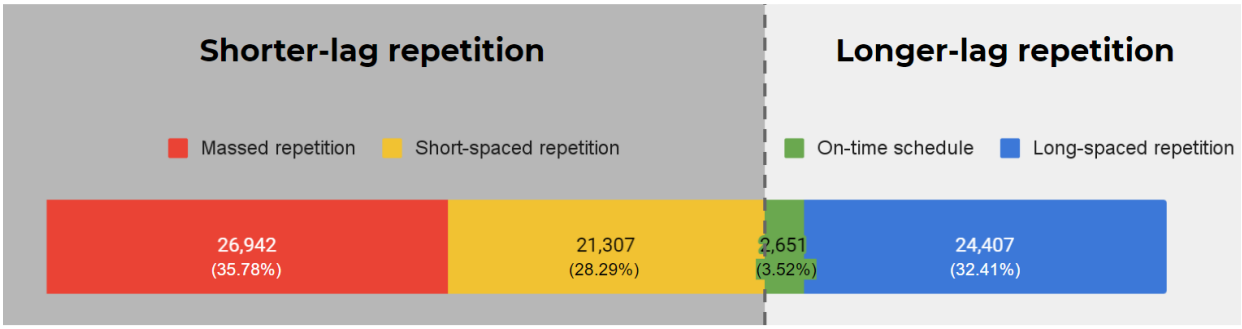


Figure 5. The number of player-question pairs following each of the four identified natural repetition patterns (massed repetition, short-spaced repetition, on-time schedule, and long-spaced repetition)

As can be seen in Figure 5, only a very small proportion of player-question pairs (2,651; 3.52%) followed the precise repetition schedule according to the Leitner Learning Box model. In contrast, the most prevalent behavior was massed repetition (26,942; 35.78%), which included

attempts made within 1 hour from the previous attempt, closely followed by long-spaced repetition (24,407; 32.41%). These results suggest that learner behaviors tended to fall into two extremes: the players first were highly engaged with the game application and crammed in a single play session, and then returned to the app to continue their learning after long periods of time.

The Effects of Players' Learning Repetition Patterns on Their Forgetting Rate

Since, in the analysis of natural repetition patterns, we observed an imbalance of player-question pairs between the four categories, in the forgetting rate analysis, we grouped the data into the following two repetition patterns (instead of four; see also Figure 6):

1) Shorter lag repetitions in which attempts to answer the questions were made before Leitner's (1972) model scheduled review time (i.e., a combination of massed repetition and short-spaced repetition);
2) Longer lag repetitions in which attempts to answer the questions were made on or after Leitner's (1972) model scheduled review time (i.e., a combination of on-schedule repetition and long-spaced repetition).

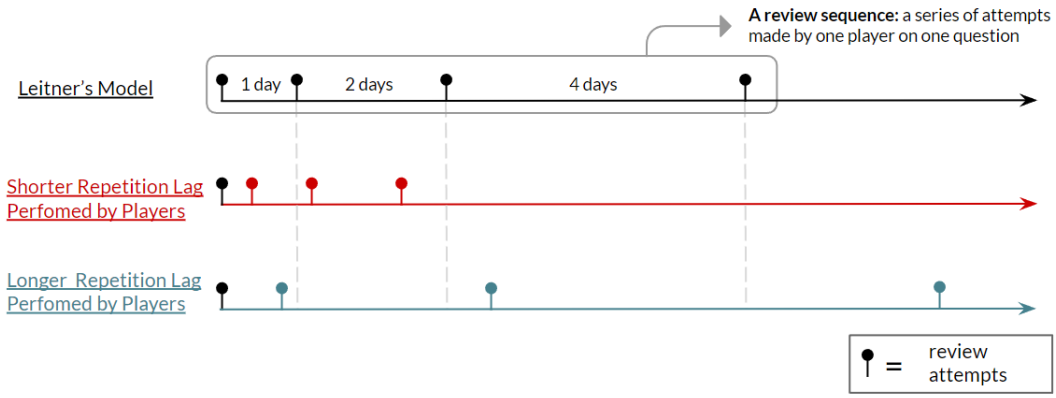


Figure 6. Examples of player-question pairs categorized as massed repetition (middle) and spaced repetition (bottom)

With these new grouping criteria, more player-question pairs (115,651) could be included in the dataset, as compared to the pairs eligible for the pattern identification analysis (75,307). This was due to the difference in the number of categories used for the pattern identification analysis and the forgetting rate analysis: four and two, respectively. With more categories in the pattern identification analysis, there was a higher likelihood that a player-question pair would be linked to more than one category. For instance, a player-question pair was included in the forgetting rate analysis in the shorter lag category if it consisted of six attempts that belonged to this category and fewer attempts in the longer lag category; however, this same pair was not included in the pattern identification analysis if the six attempts were equally divided (three attempts per category) between both massed repetition and spaced repetition.

In the forgetting rate analysis, to control for the possibility that the frequency of review performed by the player might be a contributing factor to the empirical forgetting rate, we first grouped player-question pairs based on the number of attempts that preceded attempt n (i.e., the attempt at which the empirical forgetting rate was calculated). Within each of these groups, each player-question pair was then categorized into either the shorter lag repetition group (i.e., when most attempts in the reviewing sequence were made before the scheduled review time) or the longer lag repetition group (i.e., when most attempts in the reviewing sequence were made on or after the scheduled review time). The empirical forgetting rate was then calculated for each reviewing sequence as a measure of players' knowledge retention. Figure 7 summarizes the results for reviewing sequences with up to eight attempts for each of the two repetition categories. In this analysis, we excluded player-question pairs that consisted of more than eight

attempts, as most player-question pairs were found to consist of eight or fewer attempts. As a result, the data for groups of player-question pairs that consisted of more than eight attempts were insufficient for a reliable statistical analysis. The final sample size included in the forgetting rate analysis was 114,068 player-question pairs (Pre A level = 11,322 datasets, other levels combined = 102,746 datasets). The results showed that, when players reviewed questions on or after Leitner's (1972) scheduled review time, they tended to have significantly lower forgetting rates across all groups.

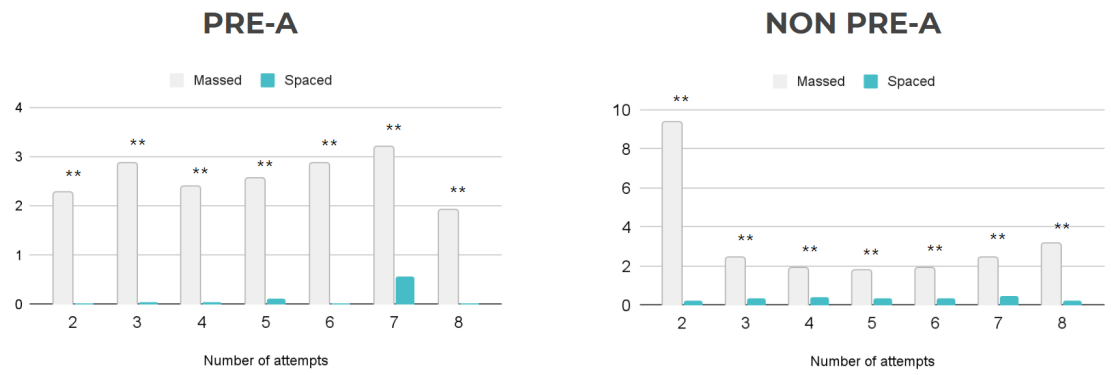


Figure 7. Median normalized empirical forgetting rates for the massed repetition group and the spaced repetition group

** $p < 0.05$ level of significance (one-tailed Mann-Whitney *U* test; Mann & Whitney, 1947).

Discussion

Factors Influencing Natural Repetition Patterns Performed by Dawn of Civilization Players

As players of Dawn of Civilization are not made explicitly aware of the QR system implemented within the game, the learning patterns identified in the present study reflect natural

interaction behaviors between players and the Dawn of Civilization game. While the EFL learners in our study were not intentionally assigned to specific learning repetition schedules (e.g., Chukharev-Hudilainen & Klepikova, 2016; Lee et al., 2021; Suzuki & Hanzawa, 2021), other factors could have influenced players' learning repetition patterns.

In this study, the learning repetition pattern for a reviewing sequence was defined as massed repetition when most attempts were made within one hour from the previous attempt. Accordingly, most attempts were likely made within one sitting (i.e., a single learning session within the game). When the study participants were immersed in the learning and gaming experience in Dawn of Civilization, they were more likely to play longer game sessions and, accordingly, consumed a lot of content in one sitting. Consequently, when a particular game ran out of ready-to-review questions, the system had to present the questions that the study participants had recently seen earlier during the session, which caused early repetition.

The other three identified repetition patterns might be related to player retention of the application where players come back to the app after leaving for longer periods of time. In our study, the study participants did not receive any reminders to return to the game (e.g., push notifications[4]), nor were they pre-assigned to a particular review schedule. Therefore, the study participants' return to Dawn of Civilization long after their last session could have been influenced by internal factors, such as learning motivation.

The reasons outlined above explain why a much smaller number of reviewing sequences was observed in Leitner's (1972) repetition schedule. Naturally, in a personalized learning environment such as a DGBL application, players cannot always follow the learning schedule, as

[4] A push notification is a message that pops up on an online platform, such as a web and a mobile application. The application publishers can send the message at any time within the application, or whenever the device is on to receive them.

they have the freedom to choose when they want to interact with the application. Due to the very narrow time range within which on-schedule repetition could occur, almost all observed behaviors were most probably due to chance. As it is unlikely that learners would perform repetition on time according to Leitner's (1972) repetition schedule, comparing forgetting rates between early or late reviewing relative to the scheduled review time was sensible for the purpose of the present investigation.

Spaced Repetition-Powered DGBL for a Reduction of Forgetting Rate

As observed in several previous studies, a DGBL application allows its players to maintain authority over their learning while retaining the potential to transform EFL learning into a playful and motivational context (Dehghanzadeh et al., 2019; Schimanke et al., 2017; Zhang et al., 2021). In Dawn of Civilization, players can interact with the application anywhere and anytime as long as they are connected to the device on which the application is installed. However, since Dawn of Civilization was successful in engaging its players to continuously come back to learn, there was a high risk of the student participants' being presented with learning materials which were not ready-to-review in the same learning session. Consistently with the results reported in several previous studies (Roediger & Butler, 2011; Roediger & Karpicke, 2006), our findings revealed that reviewing learning materials too early could compromise knowledge retention. Accordingly, whenever possible, such an early review of learning materials should be avoided whenever possible.

As discussed in the Literature Review, in EFL learning contexts, reducing the forgetting rate and promoting knowledge retention are essential for the development of learners' communication skills. Therefore, while the design and implementation of spaced repetition on a

DGBL platform should be able to get players sufficiently interested to enjoy their learning experience in the game, this involvement should also be consistently maintained over longer periods of time, rather than occur in bursts of long learning sessions. This is a simple yet extremely challenging task for product researchers and developers, as sustaining players' engagement, retention, and learning patterns require continued experimentation and iteration.

Furthermore, for effective and efficient language learning, care should be taken to prevent any possibility of premature (i.e., too early) repetition. In Dawn of Civilization, the point of reference used to define early or late reviews is Leitner's (1972) spaced repetition model, which assigns increasing interval periods. One way to avoid too early repetition is to use incentives, which have previously been reported to be a useful tool to nudge players towards performing desired behaviors (Richter et al., 2015; Bayuk & Altobello, 2019). Indeed, in-app incentives can be effectively employed to incentivize players to return and continue their learning at times where the materials are ready for review. Coupled with push notifications as reminders for players to return to the application, incentives could serve a powerful tool to maintain users' engagement with the app without compromising their learning retention (Schimanke et al., 2017).

Another possible approach to avoid too early repetition in DGBL-based applications is to ensure availability of a large pool of materials within the game—large to an extent that consuming all available learning materials in one sitting would be almost impossible. Therefore, during sessions when players are highly engaged, there is always a sufficient supply of new learning materials to present to the player without having to re-display the materials that the player has only recently been exposed to. Of note, however, this approach only minimizes the

likelihood of an early review of learning materials, rather than completely erases it. Accordingly, if complete prevention of repetition is to be achieved, this approach should be coupled with other strategies, such as the use of incentives.

In addition to what product researchers and developers can do to improve spaced repetition-powered DGBL applications, teachers can use DGBL applications to support learning. The use of DGBL applications in EFL learning can contribute to the development of learner autonomy and, consequently, lead to a reduction of teaching hours. In fact, according to Schmitt (2007), EFL learners, especially in Asia, were reported to typically need 800-1200 hours of instruction to acquire necessary vocabulary for basic daily conversation (Schmitt, 2007). Accordingly, as argued by Waluyo and Bucol (2021), teachers cannot rely solely on classroom teaching to help learners build their English lexicon to a point sufficient for active communication. Similarly, Fithriani (2021) reported that, as compared to classroom instruction alone, mobile-assisted gamification can more effectively help EFL learners to build and retain their vocabulary knowledge (see also Literature Review for further evidence). Yet, appropriate teaching strategies involving DGBL applications should also include sufficient teachers training on how to blend such applications with more conventional teaching methods to enhance learning and improve learners' knowledge retention.

Conclusion and Directions for Future Research

The present study was driven by the need to highlight the benefits of implementation of gamified EFL learning, be it at school or at home, both with teachers and alone. Based on the results, we can conclude that DGBL offers a huge potential in engaging learners and, at the same time, reducing learners' forgetting rates, which could positively impact their knowledge

retention. Knowledge retention is essential for EFL learners because it sustains learners' ability to comprehend and communicate well in a foreign language.

Taken together, the results of the present study revealed that (1) Dawn of Civilization players were unlikely to follow Leitner's (1972) repetition schedule because they were free to interact with the application anytime they wanted; and (2) those players who closely followed the game's recommended spaced repetition schedule or reviewed after the recommended schedule demonstrated significantly lower forgetting rates, indicating better knowledge retention.

The present study has several limitations. Firstly, the forgetting rate values obtained in the present study hold only within the studied context and should not be compared with the results of other studies conducted in different contexts. This lack of generalizability of our findings is associated with the way in which we adapted the use of the forgetting formula. Specifically, we used the values of 0.99 and 0.01 (instead of 1 and 0) for correct and incorrect answers, respectively. Overall, the ways in which these values are approached have a considerable impact on the empirical forgetting rate values. For instance, approaching 1 with 0.99 or 0.999 would yield very different values. Accordingly, forgetting rate values obtained in the present study should not be compared with these values obtained in previous or future studies using the same formula.

Secondly, while the our findings replicated the previously reported conclusion that, compared to massed repetition, spaced repetition is more beneficial in terms of reducing learners' forgetting rate (see Chukharev-Hudilainen & Klepikova, 2016; Seibert Hanson & Brown, 2020; Tabibian et al., 2019), we relied exclusively on Leitner's (1972) spaced repetition system with increasing time intervals between reviews. To overcome this limitation in further

research, future studies can test and compare differences in forgetting rate for various spaced repetition schedules (e.g., with equally spaced intervals or decreasing intervals).

Thirdly and finally, the present study was limited by the use of a computational approach to measure the forgetting rate. To address this limitation, future research would need to be conducted in a more controlled environment where the validity of the computational approach would be tested using a test-based approach.

References

Amalia, T. (2018). *Solve Education! Core curriculum*. https://solveeducation.org/research/se-core-curriculum/

Bayuk, J., & Altobello, S. A. (2019). Can gamification improve financial behavior? The moderating role of app expertise. *International Journal of Bank Marketing, 37*(4). 951-975. https://doi.org/10.1108/IJBM-04-2018-0086

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics, 31*(4), 635-650. https://doi.org/10.1017/S0142716410000172

Block, J. H. (Ed.). (1971). *Mastery learning: Theory and practice.* New York: Holt, Rinehart and Winston.

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research, 74*(4), 245-248. https://doi.org/10.1080/00220671.1981.10885317

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354-380. https://doi.org/10.1037/0033-2909.132.3.354

Chaikovska, O., & Zbaravska, L. (2020). The efficiency of Quizlet-based EFL vocabulary learning in preparing undergraduates for the state English exam. *Advanced Education*, *7*(14*)*, 84-90. https://doi.org/10.20535/2410-8286.197808

Chukharev-Hudilainen, E., & Klepikova, T. A. (2016). The effectiveness of computer-based spaced repetition in foreign language vocabulary instruction: A double-blind study. *Calico Journal, 33*(3), 334-354. https://journals.equinoxpub.com/CALICO/article/view/26055/26634

Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaee, E., & Noroozi, O. (2019). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning*, *34*(7), 934-957. https://doi.org/10.1080/09588221.2019.1648298

Ebbinghaus, H. (1885). *Über das Gedchtnis. Untersuchungen zur experimentellen Psychologie*. Leipzig: Duncker & Humblot; the English edition is Ebbinghaus, H. (1913). *Memory: A Contribution to Experimental Psychology*. New York: Teachers College, Columbia University.

Fithriani, R. (2021). The utilization of mobile-assisted gamification for vocabulary learning: Its efficacy and perceived benefits. *Computer Assisted Language Learning Electronic Journal (CALL-EJ), 22*(3), 146-163. http://callej.org/journal/22-3/Fithriani2021.pdf

Godwin-Jones, R. (2011). Mobile apps for language learning. *Language learning & technology, 15*(2), 2-11. http://dx.doi.org/10125/44244

Goossens, N. A., Camp, G., Verkoeijen, P. P., Tabbers, H. K., & Zwaan, R. A. (2012). Spreading the words: A spacing effect in vocabulary learning. *Journal of Cognitive Psychology, 24*(8), 965-971. https://doi.org/10.1080/20445911.2012.722617

Hung, H. T., Yang, J. C., Hwang, G. J., Chu, H. C., & Wang, C. C. (2018). A scoping review of research on digital game-based language learning. *Computers & Education, 126*, 89-104. https://doi.org/10.1016/j.compedu.2018.07.001

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. John Wiley & Sons.

Kasprowicz, R. E., Marsden, E., & Sephton, N. (2019). Investigating distribution of practice effects for the learning of foreign language verb morphology in the young learner classroom. *The Modern Language Journal, 103*(3), 580-606. https://doi.org/10.1111/modl.12586

Kickmeier-Rust, M. D., Peirce, N., Conlan, O., Schwarz, D., Verpoorten, D., & Albert, D. (2007). Immersive digital games: the interfaces for next-generation e-learning? In Stephanidis C. (eds.), *Universal Access in Human-Computer Interaction. Applications and Services*. Springer.

Kingsley, T. L., & Grabner‐Hagen, M. M. (2018). Vocabulary by gamification. *The Reading Teacher, 71*(5), 545-555. https://doi.org/10.1002/trtr.1645

Lai, C., & Zheng, D. (2018). Self-directed use of mobile devices for language learning beyond the classroom. *ReCALL, 30*(3), 299-318. https://doi.org/10.1017/S0958344017000258

Lee, I. H., Maechtle, C., & Hu, C. F. (2021). Enhancing vocabulary retention in low-achieving EFL students: Massed or spaced?. *English Teaching & Learning*, 1-16. https://doi.org/10.1007/s42321-020-00074-y

Leitner, S. (1972). *So lernt man lernen. Angewandte Lernpsychologie – ein Weg zum Erfolg*. Verlag Herder.

Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition, 40*(3), 651-675. https://doi.org/10.1017/S0272263117000341

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, *18*(1), 50-60. https://doi.org/10.1214/aoms/1177730491

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9*(5), 596-606. https://doi.org/10.1016/S0022-5371(70)80107-4

Miles, S. W. (2014). Spaced vs. massed distribution instruction for L2 grammar learning. *System, 42*, 412-428. https://doi.org/10.1016/j.system.2014.01.014

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition, 37*(4), 677-711. https://doi.org/10.1017/S0272263114000825

Nakata, T., & Suzuki, Y. (2019). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition, 41*(2), 287-311. https://doi.org/10.1017/S0272263118000219

Nushi, M., & Eqbali, M. H. (2017). Duolingo: A mobile application to assist second language learning. *Teaching English with Technology, 17*(1), 89-98. https://eric.ed.gov/?id=EJ1135889

Ortega, L. (2009). *Understanding second language acquisition*. Routledge.

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment (CIE), 1*(1), 21. https://doi.org/10.1145/950566.950596

Richter, G., Raban, D. R., & Rafaeli, S. (2015). Studying gamification: The effect of rewards and incentives on motivation. In T. Reiners & L. C. Wood (eds.), *Gamification in education and business* (pp. 21-46). Springer.

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

Rogers, J. (2015). Learning second language syntax under massed and distributed conditions. *TESOL Quarterly, 49*(4), 857-866. https://doi.org/10.1002/tesq.252

Rogers, J., & Cheung, A. (2020). Input spacing and the learning of L2 vocabulary in a classroom context. *Language Teaching Research, 24*(5), 616-641. https://doi.org/10.1177/1362168818805251

Sanatullova-Allison, E. (2014). Memory retention in second language acquisition and instruction: Insights from literature and research. *IAFOR Journal of Language Learning, 1*(1), 1-13. https://eric.ed.gov/?id=EJ1167235

Seibert Hanson, A. E., & Brown, C. M. (2020). Enhancing L2 learning through a mobile assisted spaced-repetition tool: an effective but bitter pill? *Computer Assisted Language Learning*, *33*(1-2), 133-155. https://doi.org/10.1080/09588221.2018.1552975

Schimanke, F., Mertens, R., & Vornberger, O. (2017). Designing for motivation: Design-considerations for spaced-repetition-based learning games on mobile devices. *International Journal on E-Learning, 16*(3), 287-311. Waynesville, NC USA: Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/149909/

Schmitt, N. (2007). Current perspectives on vocabulary teaching and learning. In J. Cummins & C. Davidson (eds.), *International handbook of English language teaching* (pp. 827-841). Springer.

Settles, B., & Meeder, B. (2016, August). A trainable spaced repetition model for language learning. In *Proceedings of the 54th annual meeting of the association for computational linguistics, 1*, 1848-1858. Berlin, Germany: Association for Computational Linguistics. https://10.18653/v1/P16-1174

Suzuki, Y. (2017). The optimal distribution of practice for the acquisition of L2 morphology: A conceptual replication and extension. *Language Learning, 67*(3), 512-545. https://doi.org/10.1111/lang.12236

Suzuki, Y., & DeKeyser, R. (2017). Effects of distributed practice on the proceduralization of morphology. *Language Teaching Research, 21*(2), 166-188. https://doi.org/10.1177/1362168815617334

Suzuki, Y., & Hanzawa, K. (2021). Massed task repetition is a double-edged sword for fluency development: An EFL classroom study. *Studies in Second Language Acquisition*, 1-26. https://doi.org/10.1017/S0272263121000358

Tabibian, B., Upadhyay, U., De, A., Zarezade, A., Schölkopf, B., & Gomez-Rodriguez, M. (2019). Enhancing human learning via spaced repetition optimization. *Proceedings of the National Academy of Sciences, 116*(10), 3988-3993. https://doi.org/10.1073/pnas.1815156116

Toppino, T. C., & Bloom, L. C. (2002). The spacing effect, free recall, and two-process theory: A closer look. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 437-444. https://doi.org/10.1037/0278-7393.28.3.437

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. *Psychology of learning and motivation*. *60*, 113-189. https://doi.org/10.1016/B978-0-12-800090-8.00004-4

Tu, Y., Zou, D., & Zhang, R. (2020). A comprehensive framework for designing and evaluating vocabulary learning apps from multiple perspectives. *International Journal of Mobile Learning and Organisation, 14*(3), 370-397. https://doi.org/10.1504/IJMLO.2020.108199

Waluyo, B. & Bucol, J. L. (2021). The impact of gamified vocabulary learning using Quizlet on low-proficiency students. *Computer-Assisted Language Learning Electronic Journal (CALL-EJ), 22*(1), 164-185. http://callej.org/journal/22-1/Waluyo-Bucol2021.pdf

Weissheimer, J., de Souza, J. G. M., Antunes, J. P. L., & de Souza Filho, N. S. (2019). Gamification and L2 vocabulary learning: The Vocabox experience in the languages without borders program. *Revista Linguagem & Ensino, 22*(4), 1136-1154. https://doi.org//10.15210/RLE.V22I4.16453

Wozniak, P. A. (1990). *Optimization of learning: Simulation of the learning process conducted along the SuperMemo schedule.* https://www.supermemo.com/en/archives1990-2015/english/ol

Zhang, R., Zou, D., & Xie, H. (2021). Spaced repetition for authentic mobile-assisted word learning: Nature, learner perceptions, and factors leading to positive perceptions. *Computer Assisted Language Learning*, 1-34. https://doi.org/10.1080/09588221.2021.1888752

Zhonggen, Y. (2018). Differences in serious game-aided and traditional English vocabulary acquisition. *Computers & Education, 127*, 214-232. https://doi.org/10.1016/j.compedu.2018.07.014

Zhu, Y., Fung, A. S., & Wang, H. (2012). Revisiting the effects of pronunciation and stroke order animation in digital flashcards on memorization of Chinese new words: Towards an adapted dual coding theory. *Calico Journal, 29*(3), 563-577. https://journal.equinoxpub.com/Calico/article/view/17048