Investigating the Effect of Retrieval Practice on EFL Students' Retention of Prose Passages: Repeated Testing versus Studying

Sameer Aljabri[1]

[1] Department of English, Umm Al-Qura University, Saudi Arabia

Correspondence: Sameer Aljabri, Department of English, Umm Al-Qura University, Saudi Arabia. E-mal: ssjabri@uqu.edu.sa

Received: April 2, 2024 Accepted: May 11, 2024 Online Published: June 3, 2024

doi:10.5430/wjel.v14n5p303 URL: https://doi.org/10.5430/wjel.v14n5p303

Abstract

This research aimed to investigate the beneficial effect of retrieval practice on the short- and long-term retention of prose passages by Saudi EFL speakers. A total of 270 Saudi EFL students were divided into six groups to compare the effectiveness of repeated study and repeated testing. Each student read the same 333-word passage containing 20 ideas. Results of immediate tests showed that more ideas were recalled by the repeated study (SSSS) group than by the repeated test (STTT) group, but the difference was not statistically significant. Results of delayed tests showed that the STTT group or even the single test (SSST) group was more successful with long-term retention than the SSSS group. Furthermore, the SSSS group retained less information in the first week after the study period than the other groups. The results of the study indicate that testing has a positive effect on the retention of previously learned material.

Keywords: testing effect, repeated study, repeated testing, retrieval practice, long-term retention

1. Introduction

Testing has become the central issue in most educational settings. It has been established not only as an evaluation tool but also as a potent enhancer of memory and learning (Carpenter, Pashler, & Vul, 2006). Beyond mere assessment, testing also functions as a diagnostic tool, aiding in identifying students' strengths and weaknesses (Brame & Biel, 2015). The heavy emphasis on tests nowadays obscures one of the functions of testing: the promotion of learning. Considerable research in the field of cognitive psychology has shown that utilizing multiple tests further improves knowledge retention and reduces forgetfulness compared to relying on a single test (Karpicke, 2012; Roediger & Karpicke, 2006a; for meta-analytic evidence, see Rowland, 2014). This effect is called the testing effect or retrieval practice effect. It is also known as test-enhanced learning (Yang et al., 2021).

This cognitive strategy has been extensively explored by experimental psychologists. Traditionally, this research has taken place in laboratory settings using diverse learning materials, such as word lists or pairs (Carpenter, 2009; 2011; Cranney et al., 2009; Halamish & Bjork, 2011; Peterson & Mulligan, 2013; Goossens et al., 2014; Abel & Bäuml, 2016), factual lists (Carpenter et al., 2008), face-name pairs (Carpenter & DeLosh, 2006), images (Pastötter et al., 2013), prose passages (Chan et al., 2006; Roediger & Karpicke, 2006b; Clark & Svinicki, 2014; Little & McDaniel, 2015; Bae et al., 2019), foreign language vocabulary (Kang & Pashler, 2014), brief multimedia lessons (Johnson & Mayer, 2009), PowerPoint presentations (Vojdanoska et al., 2009), and map-learning tasks (Carpenter & Pashler, 2007).

However, no published studies have explored the testing effect on EFL students in particular. This study focuses on investigating the effects of retrieval practice and repeated studying on short-term and long-term retention with Saudi undergraduate EFL students. By addressing this gap in the existing literature, the study has the potential to contribute significantly to our understanding of effective learning strategies for EFL students, particularly within the context of Saudi Arabia.

2. Literature Review

Meta-analyses conducted by Phelps (2012), Rowland (2014), Adesope et al. (2017), Pan and Rickard (2018), and Yang et al. (2021) consistently affirm that retrieval practice surpasses repeated study in facilitating long-term retention. Conversely, repeated study proves more effective for short-term retention. Roediger and Karpicke (2006b) demonstrated the testing effect using prose passages to investigate the impact of repeated study on long-term retention. Their study involved 180 undergraduate students divided into three groups: SSSS, SSST (single test), and STTT. In the SSSS group, participants studied the passage four times, each for five minutes. The SSST group read the passage three times and then underwent a recall test where they had to recall as much information as possible from the passage. The STTT group studied the passage once and then underwent three recall tests using the same format as the SSST group. Two final retention tests were administered to all groups: one immediately after the study session and the other one week later. The results indicated that additional study (SSSS) led to better performance in the immediate test, while repeated testing (STTT) improved performance in the delayed test. Specifically, in the immediate tests, the SSSS group outperformed the SSST group, which in turn outperformed the STTT group. However, in the delayed tests, the STTT group exhibited superior performance compared to the SSSS group.

Karpicke and Roediger (2007) investigated the impact of testing on multilateral free recall. Their experiment involved subjects studying a word list under three conditions: repeated study (SSSS), repeated test (STTT), and standard condition (STST). The findings revealed that increased testing led to better long-term retention, with repeated study failing to improve one-week retention rates. This study emphasizes the role of testing in both assessing and enhancing learning, emphasizing that it is the key for later retention.

Uçar and Demiraslan Çevik (2018) explored the effects of different learning conditions on middle school students. One hundred and two sixth-grade students were assigned to four groups: Study groups (SST and SSST) and Test groups (STT and STTT). Immediate and delayed test results revealed that the STTT and STT groups outperformed the SST and SSST groups in memory retention tasks.

In a laboratory setting, Bae et al. (2019) examined retrieval practice alongside four other retrieval-based study strategies: free recall, multiple-choice, test generation, and keyword. Three hundred and thirty-eight undergraduates were assigned to one of nine conditions. Subjects read a prose passage of 275 words for seven minutes and were tested one week later. The results indicated that both single and repeated retrieval conditions outperformed rehearsal conditions. Furthermore, free recall and multiple-choice emerged as the most effective strategies for retrieval practice.

Uner, Tekin, and Roediger III (2022) conducted a study on retrieval practice using true-false questions within a laboratory environment. The experiment involved ninety-one undergraduate participants who were divided into three groups: one group took true-false quizzes (True-False), another group took true-false quizzes and subsequently corrected the false statements (Correction), and a control group read several statements and then typed them into a textbox (Typing control). All participants read eight passages ranging from 533 to 605 words for five minutes each. They were tested two days after the initial phase, with eight true-false questions created for each passage. The results demonstrated that utilizing true-false questions improved retention compared to mere rereading.

Conflicting findings have emerged regarding the effects and effectiveness of various testing formats on both short-term and long-term retention. Smith and Karpicke (2014) suggest that free recall, multiple-choice (MC), and hybrid testing formats yield practically equivalent benefits under specific conditions. In contrast, Rowland (2014) asserts that short-answer questions elicit a more robust testing effect compared to multiple-choice questions. However, in another meta-analysis, Adesope et al. (2017) reach the opposite conclusion, finding that MC tests exhibit a more pronounced testing effect. Greving and Richter (2018) support the notion that short-answer questions positively impact both short-term and long-term retention.

Notably, test formats that necessitate generating an answer, such as free recall, cued recall, and short answer, tend to yield more significant testing effects than formats relying on selecting an answer, such as multiple-choice questions (Carpenter & DeLosh, 2006; Bae et al., 2019; Rawson & Zamary, 2019). This superiority of retrieval practice involving formats like cued recall over those like recognition likely arises from an elaborative or effortful process associated with generating a response.

Considerable research has explored the impact of testing versus restudy, encompassing both immediate and delayed testing scenarios. Generally, these studies indicate that test trials result in superior retention, as evaluated by long-term tests, compared to study trials. However, repeated study may sometimes enhance short-term retention. Across various investigations, research consistently underscores the efficacy of retrieval-based strategies for learning, contrasting them with less active approaches like highlighting and rereading (Adesope et al., 2017; Dunlosky et al., 2013; Karpicke, 2012; Roediger & Karpicke, 2006b; Rowland, 2014).

Recent research has delved into participants' perceptions of self-testing strategies, shedding light on how individuals recognize the benefits of these methods. Tullis and Maddox (2020) are among those who have explored this topic, revealing that students often underestimate the effectiveness of retrieval practice strategies when compared to passive rereading methods. Furthermore, studies such as those conducted by Dirkx et al. (2019) and Rivers (2020) have echoed similar findings, highlighting the tendency for individuals to undervalue the efficacy of retrieval practice strategies. This discrepancy in perception underscores the importance of raising awareness about the benefits of active learning techniques like retrieval practice. Moreover, recent research has also examined the differences between individual and collaborative retrieval practice and their respective effects on learning outcomes. Patiwael et al, (2021), Deshpande et al. (2020), Button et al. (2021), LoGiudice et al. (2023), and Knopps et al. (2023) are among the researchers who have contributed to this area of inquiry. Their studies provide insights into how collaborative retrieval practice may offer unique advantages or challenges compared to individual practice, offering valuable considerations for educators and learners alike.

The current study extends the line of inquiry by examining retrieval practice with non-native English speakers, specifically EFL (English as a Foreign Language) students. To the best of our knowledge, no published studies have explored the testing effect on Saudi students. Thus, this study seeks to fill this gap by investigating the effects of repeated testing and studying on the short-term and long-term retention of prose materials among EFL students. Understanding how retrieval practice impacts the retention of English language skills among EFL students is essential for informing teaching practices and curriculum development in this context. Additionally, examining the comparative effectiveness of retrieval practice versus repeated studying provides valuable insights into which learning approach may be more beneficial for EFL learners in the long term

In this study, free recall tests without feedback were employed to compare students' performance across three learning conditions: SSSS, SSST, and STTT. The STST condition was excluded from this experiment to maintain focus, as the feedback gained from restudying material after a test introduces an additional variable beyond the scope of this study.

3. Methodology

3.1 Participants

270 male EFL students from the Department of English at Umm Al-Qura University (UQU) in Saudi Arabia participated in the study. They were randomly selected at and allocated to one of six groups, each consisting of 45 subjects. The groups were defined based on different combinations of study and test conditions: repeated study – immediate test, repeated study – delayed test, single test – immediate test, single test – delayed test, repeated test – immediate test, and repeated test – delayed test. Details of each condition are provided in Figure 1. All participants were required to sign consent forms outlining their rights and the option to withdraw from the study at any time. Table 1 presents a breakdown of the conditions utilized in the study.

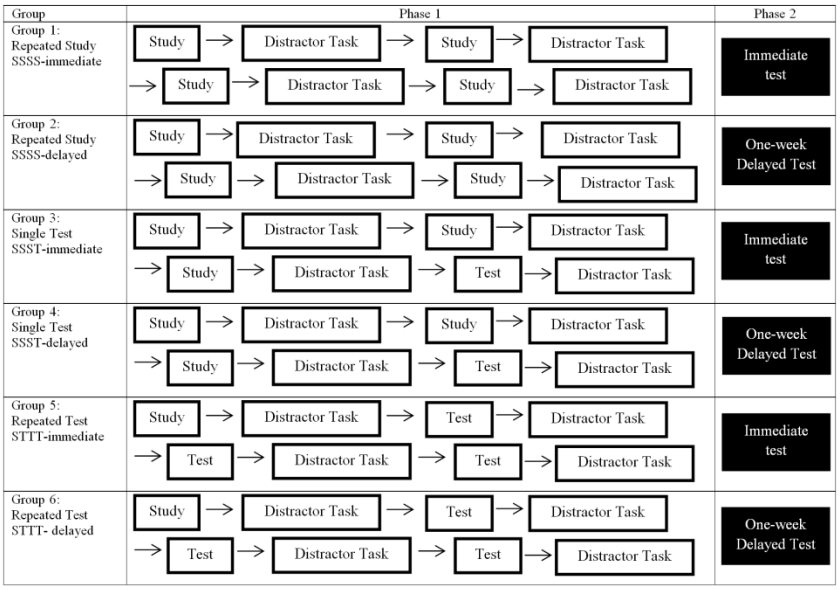

Figure 1. Learning Conditions in the Study

3.2 Materials

The study utilized a 333-word prose passage focusing on Thomas Edison, chosen by two English professors. The passage was then analyzed to identify twenty T-units, as defined by Richards, Platt, and Weber (1985), representing distinct idea units within the text.

3.3 Procedure

In Phase 1, subjects assigned to the SSSS condition were directed to read the passage for five minutes across four consecutive study sessions. Following each session, participants engaged in a two-minute distractor task involving a mathematical problem. After the final study period, they spent five minutes solving a similar mathematical problem.

Participants in the SSST group followed a similar procedure, but with three study periods instead of four. They also completed a two-minute distractor task between each session. Upon completing the study phase, subjects in this group took a free recall test, tasked with writing down as many ideas as possible from the passage. Subsequently, they solved a five-minute mathematical problem.

In the STTT condition, subjects read the passage for five minutes, solved a two-minute mathematical problem, and underwent three testing sessions. A two-minute mathematical problem followed each test, and after the final test, participants engaged in a five-minute mathematical problem-solving task.

The timeline for Phase 2 was determined by the testing interval (immediate or delayed) assigned to the subject groups. Subjects in the immediate test condition underwent a final free recall test immediately after Phase 1, while those in the one-week condition took the test a week later. During these tests, participants were instructed to recall as many ideas as possible from the passage, without regard for chronological order or exact wording.

3.4 Data Analysis

The study employed a 3 × 2 between-subject design to explore the impacts of three study conditions (SSSS, SSST, and STTT) during Phase 1, and two final test intervals (immediate and delayed) during Phase 2 on information retention. The prose passage provided to participants contained 20 distinct items of information for recall. Following the study sessions, a final test was administered either after five minutes or one week, with scores ranging from 0 to 20.The independent variables in the study were the study condition and final test interval, while the dependent variable was the score attained on the final recall test (ranging from 0 to 20).To examine the effects of the study condition and final test interval on the proportion of the 20 possible items recalled by participants, a logistic regression was conducted. Additionally, an analysis of deviance (ANOVA) was performed to assess the overall significance of the main effects of the

study condition and testing interval, as well as their interaction.

4. Results

As presented in Table 1, participants in the SSSS group demonstrated a higher recall rate (25%) compared to those in the STTT group (23.5%) in the immediate test condition. Conversely, the STTT group exhibited a higher recall rate than the SSST group (13.5%). In the delayed test condition, participants in the STTT group showed a higher recall rate (19.65%) than those in the SSST group (16.65%) and the SSSS group (7.35%).The discrepancy between the performance of the SSSS group in the immediate test (25%) and the delayed test (7.3%) was significant, indicating that retention observed in the immediate test did not carry over to the delayed test. Conversely, the difference in retention between the two STTT groups (immediate and delayed test) was insignificant (3.85% difference).Among the immediate test groups, the SSST group exhibited the lowest recall rate (13.5%), whereas among the delayed test groups, the lowest recall rate was observed in the SSSS group (7.3%).

Table 1. Mean and Standard Deviation of T-units for Study Groups

Condition	N	Mean	Std. Deviation	Std. Error of Mean
SSSS – immediate	45	5 (25%)	2.88	0.76
SSSS – delayed	45	1.46 (7.3%)	1.68	0.45
SSST – immediate	45	2.07 (13.5%)	1.86	0.49
SSST – delayed	45	3.33 (16.65%)	2.23	0.59
STTT – immediate	45	4.07 (23.5%)	2.89	0.77
STTT – delayed	45	3.93 (19.65%)	1.71	0.45

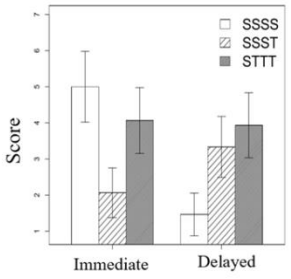

Figure 2. The mean values for each study condition and test interval combination

Figure 2 illustrates the number of ideas recalled on the final tests for each group, categorized by testing interval (immediate and delayed). Among the immediate test conditions, subjects in the SSST condition demonstrated the lowest number of recalled ideas. Conversely, among the delayed test conditions, subjects in the SSSS condition recalled the lowest number of ideas.

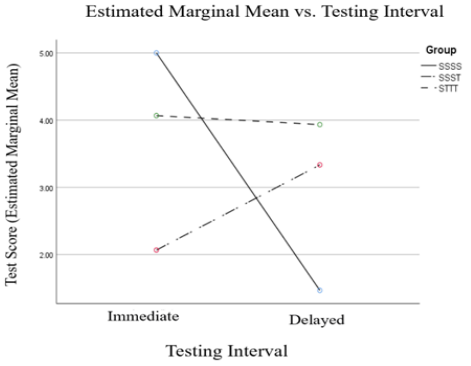

Figure 3. The estimated marginal mean for the final test for each study condition.

Figure 3 highlights a significant difference in the mean testing score between the immediate and delayed SSSS groups. However, this discrepancy was not observed between the two STTT groups

Table 2. Analysis of Variance: Between-Subjects Effect

	Sum of Squares	df	Mean Square	F ratio	Sig.
Between Groups	394.267	5	78.853	15.289	.000
Within Groups	1361.600	264	5.158		
Total	1755.867	269			

R Squared = .225; Adjusted R Squared = .178;. df = degrees of freedom; sig. = significance.

Table 2 indicates an F-value of 15.289 with 5 and 264 degrees of freedom, which is statistically significant at the 0.001 level ($P < 0.05$). The ETA Square value of 0.22 suggests that approximately 22% of the variation in accurately recalled ideas from the prose material can be explained by the differences between the six groups. Although this relationship is weak, it is statistically significant. The null hypothesis posited for this study, which states that there will be no difference between the six study conditions, can be rejected based on the results presented in the ANOVA table (Table 2).

Table 3. Post Hoc Analysis: Multiple Comparisons

LSD

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
SSSS - immediate	SSSS - delayed	3.5333 *	.84878	0.000	1.8454	5.2212
	SSST - immediate	2.9333 *	.84878	0.001	1.2454	4.6212
	SSST - delayed	1.6667	.84878	0.053	-0.0212	3.3546
	STTT - immediate	0.9333	.84878	0.275	-0.7546	2.6212
	STTT - delayed	1.0667	.84878	0.212	-0.6212	2.7546
SSSS - delayed	SSSS - immediate	-3.5333 *	.84878	0.000	-5.2212	-1.8454
	SSST - immediate	-0.6000	.84878	0.482	-2.2879	1.0879
	SSST - delayed	-1.8667 *	.84878	0.031	-3.5546	-0.1788
	STTT - immediate	-2.6000 *	.84878	0.003	-4.2879	-0.9121
	STTT - delayed	-2.4667 *	.84878	0.005	-4.1546	-0.7788
SSST – immediate	SSSS - immediate	-2.9333 *	.84878	0.001	-4.6212	-1.2454
	SSSS - delayed	0.6000	.84878	0.482	-1.087	2.2879
	SSST - delayed	-1.2667	.84878	0.139	-2.954	0.4212
	STTT - immediate	-2.000 *	.84878	0.021	-3.687	-0.3121
	STTT - delayed	-1.8667 *	.84878	0.031	-3.554	-0.1788
SSST – delayed	SSSS - immediate	-1.6667	.84878	0.053	-3.355	0.0212
	SSSS - delayed	1.8667 *	.84878	0.031	0.178	3.5546
	SSST - immediate	1.2667	.84878	0.139	-0.421	2.9546
	STTT - immediate	-0.7333	.84878	0.390	-2.421	0.9546
	STTT - delayed	-0.6000	.84878	0.482	-2.621	1.0879
STTT – immediate	SSSS - immediate	-0.9333	.84878	0.275	-2.621	0.7546
	SSSS - delayed	2.6000 *	.84878	0.003	0.912	4.2879
	SSST - immediate	2.000 *	.84878	0.021	0.312	3.6879
	SSST - delayed	0.7333	.84878	0.390	-0.954	2.4212
	STTT - delayed	0.1333	.84878	0.876	-1.554	1.8212
STTT – delayed	SSSS - immediate	-1.0667	.84878	0.212	-2.754	0.6212
	SSSS - delayed	2.4667 *	.84878	0.005	0.778	4.1546
	SSST - immediate	1.8667 *	.84878	0.031	0.178	3.5546
	SSST - delayed	0.6000	.84878	0.482	-1.087	2.2879
	STTT - immediate	-0.1333	.84878	0.876	-1.821	1.5546

Dependent Variable: Score. * The mean difference is significant at the .05 level.

Table 3 highlights several significant findings regarding the performance differences among the study conditions. First, the difference in performance between the SSSS – immediate and SSST – immediate groups is statistically significant ($p = 0.001$). Specifically, subjects in the SSSS group recalled, on average, three more items than those in the SSST group when tested immediately after the study period. Additionally, subjects in the SSSS–immediate group recalled approximately 3.5 items more than the SSSS–delayed group, indicating a significant difference associated with the mean recalled ideas between the two SSSS conditions ($P= 0.000$).

Second, although subjects in the SSSS condition recalled more ideas than those in the STTT condition during the immediate test, the difference was not statistically significant ($p = 0.275$).

Third, subjects in the SSSS – delayed group recalled around two fewer items, on average, than those in the SSST condition ($p = 0.031$), and approximately 2.5 fewer items, on average, than those in the STTT – delayed condition ($p = 0.005$). This indicates less retention in the SSSS – delayed condition compared to all other conditions.

Fourth, testing time did not significantly affect the scores for subjects in the SSST ($p = 0.139$) and the STTT ($p = 0.876$) conditions, as subjects in both groups demonstrated similar levels of idea recall during delayed testing.

Fifth, Subjects assigned to the SSST immediate group recalled an average of 2 items less than those in the STTT immediate group, and this difference was statistically significant ($p = 0.021$). However, in the delayed test, both conditions (SSST and STTT) were associated with similar levels of item recall, and this difference was not statistically significant ($p = 0.482$).

5. Discussion

The primary objective of this study was to examine the effects of repeated testing and repeated study on both short-term and long-term retention within a laboratory setting. Differing from previous research, the focus was on investigating retrieval practice among Saudi EFL participants.

The primary findings of this research reveal that while repeated study marginally enhances subjects' performance on immediate tests compared to repeated testing (25% versus 23.5%), this difference was insignificant. However, repeated testing significantly outperforms repeated study or a single test in promoting long-term retention. Subjects in the repeated testing group recalled substantially more information during delayed tests compared to those in the repeated study group (19.65% versus 7.3%). Subjects in the repeated testing group had only one study period, whereas those in the repeated study group had four. These results align with previous laboratory studies, emphasizing the superiority of repeated testing over repeated study for fostering long-term retention (Karpicke & Roediger, 2007; Roediger & Karpicke, 2006b; Johnson & Mayer, 2009; Wheeler et al., 2003).

In contrast to certain previous studies (Wheeler et al., 2003; Roediger & Karpicke, 2006b), the current study revealed no statistical superiority in performance among subjects in the repeated study group compared to those in the repeated testing group. This finding suggests that frequent practice of repeated testing can indeed enhance short-term retention among EFL students. Participants in Saudi Arabia, where this study was conducted, typically adhere to learning and studying methods predominantly reliant on repeated studying.

The study's findings reveal that repeated study yielded superior outcomes for immediate tests compared to delayed tests. This observation can be elucidated by the concept of the forgetting curve, suggesting rapid forgetting during the initial week. Consequently, the statistically significant differences in mean recalled ideas between the repeated study condition and other conditions emerged. Notably, the presence of testing was associated with a lower forgetting rate, aligning with previous research indicating reductions in forgetting rates in testing environments (Carpenter et al., 2008; Wheeler et al., 2003).

The observed enhancement in long-term retention performance can be attributed to overlapping theoretical frameworks. One such framework is the transfer-appropriate processing hypothesis, which underscores the importance of congruence between encoding and retrieval processes. According to this hypothesis, memory performance benefits when the cognitive processes engaged during encoding or studying align with those during retrieval. This alignment facilitates enhanced retention (Morris et al., 1977; Roediger, 1990; Thomas & McDaniel, 2007; Butler & Roediger, 2007; Wartenweiler, 2011).

In the current study, subjects engaged in retrieval processes during the initial test, accessing stored information from memory. This process of retrieval during the immediate test facilitated performance on subsequent delayed tests. Within the repeated testing group, subjects were compelled to repeatedly engage in retrieval practice as they recalled elements of the prose passage a week later. This aligns with the transfer-appropriate processing hypothesis, as the congruence between retrieval practice during the immediate and delayed tests resulted in improved performance. This hypothesis was substantiated by the repeated testing group's superior performance over the repeated study group in the delayed test. Another insightful theory that sheds light on the findings of our study is the concept of desirable difficulty, which was introduced by Robert Bjork (1994). This theory posits that learning tasks requiring an optimal level of challenge can significantly enhance long-term retention. According to this framework, administering a test after the initial encoding phase is more conducive to retention compared to restudying. This is because testing entails cognitive engagement, elaboration, and manipulation of information, whereas restudying often involves passive review and superficial processing.

In essence, testing prompts learners to actively engage with the material, fostering deeper understanding and memory consolidation. Conversely, restudying may lead to poorer long-term retention as it lacks the desirable level of cognitive effort associated with testing (Cranney et al., 2009). Empirical evidence from various studies supports this hypothesis, underscoring the superiority of testing over restudying in facilitating long-term retention (Roediger & Karpicke, 2006a; Greving & Richter, 2018; Stenlund et al., 2016). The results of our current study align with the concept of desirable difficulty, as repeated retrieval practice was found to be associated with enhanced retention of the prose passage. This enhancement can be attributed to the cognitive challenges posed by delayed testing, which encouraged deeper engagement with the material and facilitated robust memory formation. Thus, our findings underscore the efficacy of testing as a strategy for promoting long-term retention, in accordance with the principles of desirable difficulty.

While Roediger and Karpicke (2006b) found that subjects in the SSST group performed better on the immediate test compared to the SSST–delayed group, our study yielded contrasting results. In our investigation, although there was a slight increase in the SSST group's mean score on the immediate test, this difference was not statistically significant. Moreover, this marginal improvement did not impact the

overall efficacy of repeated testing on long-term retention. One possible explanation for this discrepancy is the presence of individual differences among participants in our study. Factors such as language comprehension, background knowledge, and metacognitive abilities could have influenced subjects' performance across different testing conditions. These individual variations might have mitigated the observed differences between the SSST and SSST–delayed groups, leading to non-significant differences in performance. It's essential to acknowledge the nuanced interplay of these factors in shaping learning outcomes and to consider them when interpreting the results of our study. While our findings may diverge from previous research in some aspects, they provide valuable insights into the complex dynamics of retrieval practice and its impact on long-term retention in laboratory contexts.

An additional noteworthy finding from this study is the significant impact of repeated testing, or even a single test, on long-term retention compared to repeated studying. The performance differences observed on the delayed test strongly favored the SSST and STTT groups, indicating their superior retention over time. These findings underscore the importance of integrating testing not solely as an evaluative tool but also as a potent teaching method in educational practices. Teachers are encouraged to recognize the efficacy of tests as a means to enhance learning outcomes. By incorporating regular testing into instructional strategies, educators can facilitate deeper understanding, reinforce memory retention, and promote meaningful learning experiences for their students.

6. Conclusion

This study aimed to investigate the impact of retrieval practice on learning, specifically focusing on non-native English speakers, with data collected from 270 Saudi Arabian EFL students. The effects of repeated testing and repeated study on short-term and long-term retention of prose passages were examined. Participants were randomly assigned to six groups (SSSS-immediate, SSSS-delayed, SSST-immediate, SSST-delayed, STTT-immediate, STTT-delayed), and each read a 333-word prose passage containing 20 ideas. Tests were administered immediately after the study phase (immediate groups) and one week later (delayed groups). The findings suggest that repeated testing is more effective than repeated study for long-term retention, although repeated testing may be less effective for immediate tests. The results support the notion that testing reduces the rate of forgetting among students, indicating a positive impact of testing on the retention of learned materials. These findings have significant implications for classroom practice, emphasizing the importance of using tests not only for assessment but also as a tool to support the teaching process. The findings of this study can potentially inform educators and policymakers in Saudi Arabia and beyond about the most effective strategies for promoting English language learning and retention among undergraduate EFL students. Moreover, by filling this gap in the literature, the study contributes to the broader conversation on evidence-based teaching practices and the optimization of learning outcomes in diverse educational contexts. This study, conducted with male Saudi Arabian university students in a laboratory context, serves as a foundation for future research on the effects of testing on both male and female learners' short-term and long-term retention. Future studies could explore various testing formats, such as cued recall and recognition tests, and investigate the role of feedback in repeated testing and repeated study. Additionally, extending the testing intervals beyond one week to assess retention over longer periods, such as one or three months, would provide further insights into the effectiveness of retrieval practice.

Acknowledgments

Not applicable

Authors contributions

Not applicable

Funding

Not applicable

Competing interests

Not applicable

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal’s policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Abel, M., & Bäuml, K. (2016). Retrieval practice can eliminate list method directed forgetting. *Memory & Cognition, 44*, 15-23. https://doi.org /10.3758/s13421-015-0539-x

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659-701. https://doi.org /10.3102/0034654316689306

Bae, C.L., Therriault, D.J. & Redifer, J.L. (2019). Investigating the testing effect: Retrieval as a characteristic of effective study strategies. *Learning and Instruction, 60*(1), 206-214. https://doi.org/10.1016/j.learninstruc.2017.12.008

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about Knowing*. (pp. 185–205). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/4561.003.0011

Brame, C. J., & Biel, R. (2015). Test-enhanced learning: the potential for testing to promote greater learning in undergraduate science courses. *CBE-Life Sciences Education, 14*(2), 1-12. https://doi.org /10.1187/cbe.14-11-0208

Butler, A. C., & Roediger, H. L., III. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology 19*, 514-527. https://doi.org /10.1080/09541440701326097

Button, P., Collins, L. A., Denteh, A., Garc´ıa-P´erez, M., Harrell, B., Isaac, E., & Ziedan, E. (2021). Teaching controversial and contemporary topics in economics using a jigsaw literature review activity. *The Journal of Economic Education*, *52*(4), 286–295. https://doi.org/10.1080/00220485.2021.1963373

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1563–1569.https://doi.org /10.1037/a0017021

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1547–1552. https://doi.org /10.1037/a0024140

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*(2), 268-276. https://doi.org/10.3758/BF03193405

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review, 14*(3), 474–478. https://doi.org /10.3758/BF03194092

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review, 13*(5), 826-830. https://doi.org /10.3758/BF03194004

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*(2), 438-448. https://doi.org /10.3758/MC.36.2.438

Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135*(4), 553–571. https://doi.org /10.1037/0096-3445.135.4.553

Clark, D. A., & Svinicki, M. D. (2014). The Effect of retrieval on post-task enjoyment of studying. *Educational Psychology Review, 27*(1), 51-67. https://doi.org /10.1007/s10648-014- 9272-4

Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology, 21*(6), 919-940. https://doi.org /10.1080/09541440802413505

Deshpande, S., Ritzenthaler, D., Sun, A., Rudert, N., & Lewis, J. (2020). A unique flipped classroom approach shows promising results in physician assistant education. *Medical Teacher, 42*(3), 285–290. https://doi.org/10.1080/0142159X.2019.1679360

Dirkx, K. J. H., Camp, G., Kester, L., & Kirschner, P. A. (2019). Do secondary school students make use of effective study strategies when they study on their own? *Applied Cognitive Psychology, 33*(5), 952-957. https://doi.org/10.1002/acp.3584

Dunlosky, J. Rawson, K. Marsh, E. Nathan, M. & Willingham, D. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions from Cognitive and Educational Psychology. *Psychol Sci Public Interest, 14*(1),4-58. https://doi.org / 10.1177/1529100612453266.

Goossens, N.A.M.C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K., & Zwaan, R.A. (2014). The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning. *Journal of Applied Research in Memory and Cognition, 3*(3), 177-182. https://doi.org/10.1037/h0101800

Greving, S., & Richter, T. (2018). Examining the testing effect in university teaching: Retrievability and question format matter. *Frontiers in Psychology*, *4*(9), Article 2412. https://doi.org /10.3389/fpsyg.2018.02412

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? a distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(4), 801-812. https://doi.org /10.1037/a0023219

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology, 101*(3), 621-629. https://doi.org /10.1037/a0015183

Kang, S. H. K., & Pashler, H. (2014). Is the benefit of retrieval practice modulated by motivation? *Journal of Applied Research in Memory and Cognition, 3*(3), 183- 188. https://doi.org /10.1016/j.jarmac.2014.05.006

Karpicke, J. D. (2012). Retrieval-Based Learning: Active Retrieval Promotes Meaningful Learning. Current Directions in *Psychological Science, 21*(3), 157-163. https://doi.org /10.1177/0963721412443552

Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 704-719. https://doi.org /10.1037/0278-7393.33.4.704

Knopps, A. G., & Wissman, K. T. (2023). Collaborative retrieval practice reduces mind-wandering during learning. *Experimental Psychology, 70*(4), 241–248. https://doi.org/10.1027/1618-3169/a000589

Little, J. L., & McDaniel, M. A. (2015). Individual differences in category learning: memorization versus rule abstraction. *Memory & Cognition*, *43*(2), 283-297. https://doi.org /10.3758/s13421-014-0475-1

LoGiudice, A. B., Heisz, J. J., & Kim, J. A. (2023). Does collaborative testing in the classroom enhance delayed transfer of knowledge? *Scholarship of Teaching and Learning in Psychology, 9*(3), 312–325. https://doi.org/10.1037/stl0000267

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behaviour, 16*(5), 519-533.https://doi.org /10.1016/S0022-5371(77)80016-9

Pan, S.C., & Rickard, T.C. (2018). Transfer of test-enhanced learning: meta-analytic review and synthesis. Psychological *Bulletin, 144*, 710-756. https://doi.org /10.1037/bul0000151

Pastötter, B., Weber, J., & Bäuml, K.-H.T. (2013). Using testing to improve learning after severe traumatic brain injury. *Neuropsychology, 27*(2), 280-285. https://doi.org /10.1037/a0031797

Patiwael, J.A., Douma, A.H., Bezâková, N., Kusurkar, R.A., & Daelmans, H.E. (2021). Collaborative testing in physical examination skills training and the autonomous motivation of students: a qualitative study. *BMC Medical Education, 21*. https://doi.org/10.1186/s12909-021-02618-7

Peterson, D. J., & Mulligan, N. W. (2013). The negative testing effect and multifactor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(4), 1287-1293.https://doi.org /10.1037/a0031337

Phelps, R. P. (2012). The effect of testing on student achievement, 1910–2010. *International Journal of Testing*, *12*(1), 21–43. https://doi.org /10.1080/15305058.2011.602920

Rawson, K. & Zamary, A. (2019). Why is free recall practice more effective than recognition practice for enhancing memory? Evaluating the relational processing hypothesis. *Journal of Memory and Language, 105*, 141-152. https://doi.org/10.1016/j.jml.2019.01.002.

Richards, J. C., Platt, J., & Weber, H. (1985). *Longman dictionary of applied linguistics*. London: Longman.

Rivers, M. L. (2021). Metacognition about practice testing: A review of learners' beliefs, monitoring, and control of test-enhanced learning. *Educational Psychology Review, 33*(3), 823– 862. https://doi.org/10.1007/s10648-020-09578-2

Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181-210. https://doi.org /10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., III, & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249-255. https://doi.org /10.1111/j.1467-9280.2006.01693.x

Roediger, H.L., III. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45*(9), 1043–1056. https://doi.org

/10.1037/0003-066x.45.9.1043

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org /10.1037/a0037559

Smith, M. A. & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory, 22* (7), 784-802. https://doi.org/10.1080/09658211.2013.831454

Stenlund, T., Sundström, A., & Jonsson, B. (2016). Effects of repeated testing on short-and long-term memory performance across different test formats. *Educational Psychology. 36*(10), 1710–1727. https://doi.org /10.1080/01443410.2014.953037

Thomas, A. K., & McDaniel, M.A. (2007). *The negative cascade* of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition, 35*(4), 668-678. https://doi.org /10.3758/BF03193305

Tullis, J.G., & Maddox, G.B. (2020). Self-reported use of retrieval practice varies across age and domain. *Metacognition and Learning, 15*, 129-154. https://doi.org/10.1007/s11409-020-09223-x.

Uçar, B., & Demiraslan Çevik, Y. (2018). Examination of Different Learning Conditions in the Testing Effect: Example of Safe Internet Use. *Bartın University Journal of Faculty of Education*, *7*(1), 29-66. https://doi.org /10.14686/buefad.322792

Uner, O., Tekin, E., & Roediger, H. L. III. (2022). True–false tests enhance retention relative to rereading. *Journal of Experimental Psychology: Applied, 28*(1), 114–129. https://doi.org /10.1037/xap0000363

Vojdanoska, M., Cranney, J., & Newell, B. R. (2009). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology, 24*(8), 1183-1195. https://doi.org /10.1002/acp.1630

Wartenweiler D. (2011). Testing effect for visual-symbolic material: enhancing the learning of Filipino children of low socio-economic status in the public school system. *The International Journal of Research and Review, 6*(1), 74–93.

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*(6), 571-580. https://doi.org /10.1080/09658210244000414

Yang, C., Luo, L., Vadilo, M, A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boots classroom learning: A systematic and meta-analytic review. *Psychological Bulletin, 147*, 399-435. https://doi.org /10.1037/bul0000309