

Claremont Colleges

## Scholarship @ Claremont

---

Scripps Senior Theses

Scripps Student Scholarship

---

2020

### Merde, Merde, Merde: The Testing Effect, Foreign Language Anxiety, and Their Impact on Foreign Language Learning

Priya Canzius

Follow this and additional works at: [https://scholarship.claremont.edu/scripps\\_theses](https://scholarship.claremont.edu/scripps_theses)

 Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), and the [Social and Behavioral Sciences Commons](#)

---

#### Recommended Citation

Canzius, Priya, "Merde, Merde, Merde: The Testing Effect, Foreign Language Anxiety, and Their Impact on Foreign Language Learning" (2020). *Scripps Senior Theses*. 1599.  
[https://scholarship.claremont.edu/scripps\\_theses/1599](https://scholarship.claremont.edu/scripps_theses/1599)

This Open Access Senior Thesis is brought to you for free and open access by the Scripps Student Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in Scripps Senior Theses by an authorized administrator of Scholarship @ Claremont. For more information, please contact [scholarship@cuc.claremont.edu](mailto:scholarship@cuc.claremont.edu).

***MERDE, MERDE, MERDE:***  
**THE TESTING EFFECT, FOREIGN LANGUAGE ANXIETY, AND THEIR**  
**IMPACT ON FOREIGN LANGUAGE LEARNING**

by  
**PRIYA CANZIUS**

**SUBMITTED TO SCRIPPS COLLEGE IN PARTIAL FULFILLMENT OF THE**  
**DEGREE OF BACHELOR OF THE ARTS**

**PROFESSOR LISE ABRAMS**  
**PROFESSOR MARINO FORLINO**

Abstract

This experiment examined the effect of foreign language classroom anxiety and assigned learning condition (repeated studying or repeated testing) on participant recall of newly-learned foreign language vocabulary word pairs. Previous research has identified a testing effect, which is a phenomenon where repeated testing is better for a learner's long-term recall than repeated studying. The hypothesis for this experiment was that the testing effect would occur for students with lower levels of foreign language classroom anxiety but not for students with higher levels of foreign language classroom anxiety. Sixteen participants recruited from Introductory French and Introductory Arabic classrooms at the Claremont Colleges were divided into two learning condition groups, study (SSST) or test (STTT). Participants were also categorized as higher or lower anxiety based on their responses to the Foreign Language Classroom Anxiety Scale. All participants studied 40 English-foreign language word pairs in Session 1. In Sessions 2 and 3, participants in the SSST condition restudied the word pairs, while participants in the STTT condition took a cued recall test where they attempted to retrieve the foreign language word in response to the English word cue. In Session 4, all participants took a cued recall test. Contrary to the hypothesis, the results showed that the testing effect occurred for participants with higher anxiety but not for participants with lower anxiety. The results of this experiment suggest that repeated testing counteracted the negative effects of anxiety on test performance for higher anxiety foreign language learners. This finding has implications for the educational strategies that should be implemented to lessen students' foreign language classroom anxiety and potentially increase their recall of foreign language vocabulary in an introductory foreign language classroom at the college level.

## The Testing Effect, Foreign Language Anxiety, and their Impact on Foreign Language Learning

Many educators see the testing effect as a reliable justification for providing their students with tests and quizzes prior to cumulative final exams. The testing effect, first introduced by scientist Edwina Abbott in 1909, states simply that testing participants on material increases their retention of said material in the future (Pastötter & Bäuml, 2014). This theory has proven to be effective in educational contexts, ranging from subjects such as science, reading comprehension and foreign language learning. Because of its association with memory retention, educators are likely to implement test-enhanced learning in their classrooms. However effective the testing effect may be, test-enhanced learning has also proven to be detrimental to students with high levels of test anxiety (Cassady, 2004). Moreover, in foreign language learning classrooms, a new type of anxiety is introduced: foreign language classroom anxiety (Horwitz, Horwitz, & Cope, 1986). The effects that repeated testing, in addition to foreign language classroom anxiety, have on memory retention and test performance have not yet been researched. Therefore, this study seeks to determine the effectiveness of the testing effect in a specific anxiety-inducing situation, i.e., foreign language learning, and whether students enrolled in a foreign language class will benefit from test-enhanced learning if they have high anxiety about learning a foreign language.

### *The Testing Effect*

In their seminal article, Roediger and Karpicke (2006) demonstrated how the testing effect can and should be applied in educational settings as a tool to improve scores on exams in addition to future retention of material. For this experiment, researchers had participants

study short passages (“The Sun” or “Sea Otters”): in Experiment 1, participants either read the passage and immediately took a test on it (testing condition) or studied it again (study condition) before they took the final test either 5 minutes, 2 days, or 1 week later. In the case of short-term retention (5 minutes), repeated studying proved to be more effective than testing. Yet, in the case of both long-term retention intervals, students who were tested prior to the final test performed better than those who were not. In Experiment 2, participants studied the passage, and either consecutively took three exams about the material (study, test, test, test; STTT), studied the passage two more times and took one test (study, study, study, test; SSST), or studied the passage three more times (study, study, study, study; SSSS) before taking the final test either 5 minutes or 1 week later. The results of Experiment 2 showed that massed practice (SSSS) was more effective for retention in the short term, but testing, especially repeated testing (STTT), helped greatly with retention in longer-term scenarios, even more than a single test session (SSST).

In an educational setting, some forms of quizzing exhibit the benefits of repeated testing more than others. McDaniel, Anderson, Derbish, and Morrisette (2007) performed an experiment in an online course to better understand the generalizability of the repeated testing within an educational setting. Participants within this course took weekly quizzes (including multiple choice and short answer questions) in addition to two unit tests and a final exam, none of which counted for their course grade. Students in the control group were exposed only to the descriptive text, which did not include quizzing. The questions presented to students were worded differently in the quizzes than they were on the unit and final tests. In this experiment, the final test was administered approximately three months after students completed the first quiz. McDaniel et al. (2007) found that the testing effect

was present for short answer questions but not for multiple choice questions and that students in the short answer question group also outperformed students in the control group. They proposed that the testing effect occurred in the short answer questions because short answer questions most resemble recall of material while multiple choice questions and reading-only merely required the recognition of facts. According to McDaniel et al. (2007), the testing effect is contingent on successful retrieval practice (in this scenario, of facts), which occur most when short answers are required of participants.

Feedback also plays an important role in testing and in the testing effect. Jang and Marshall (2018) assigned participants to four conditions based on feedback type: “(1) feedback displaying the original question and four alternative options, including the correct answer (Feedback 1); (2) feedback displaying the original question and the correct answer (Feedback 2); (3) feedback displaying only the correct answer (Feedback 3); and (4) no feedback as the control condition” (Jang & Marshall 2018, p. 111). Participants underwent a multiple-choice test of 100 general knowledge questions, which, depending on the condition, was followed by feedback. Participants returned after two days to complete a final test, which was composed of the same questions that they had been tested on the previous time. From the results of the final test, researchers learned that participants who received feedback performed better on the final test compared to participants in the control condition. However, there was no significant difference between the results of the participants in the feedback conditions. This suggests that feedback, regardless of the way it is given to a participant, is beneficial for the retention of correct answers. Jang and Marshall (2018) speculate that this is the case because “when the correct answer [feedback] is made available, people can integrate that information into memory and improve performance”

(Jang & Marshall 2018, p. 119). Feedback is an important component of testing in an educational environment, and the results of this study suggest that students will perform even better on a final test if they are provided with feedback prior to the final test.

*Foreign Language Learning and the Testing Effect*

Research has also consistently validated the testing effect in foreign language learning studies, especially when feedback is presented. One such study is that of Szöllősi, Keresztes, Novák, Szászi, Kéri, and Racsmány (2017). In their study, Szöllősi et al. (2017) examined the validity of the testing effect when the final test occurred in a stressful situation. Participants were exposed to and studied 40 Swahili-Hungarian word pairs. After the initial study of the vocabulary words, participants either restudied or were tested on the material (practice cycle) by typing in the Hungarian word equivalents of Swahili words. Participants in the testing condition were given feedback (the correct answer) for each word pair during the practice cycle. All participants returned to the lab seven days after the initial study phase for a final test, during which participants were exposed to either a stressful or non-stressful (control) task and their stress levels were assessed before and after this task. In the stress condition, participants underwent the Trier Social Stress Task (TSST), which included a 5-minute speech and a 5-minute verbal arithmetic assessment in an effort to induce anxiety. Participants in the control condition performed a control TSST, which did not contain the stress-inducing elements of the TSST in the stress condition but was created to reflect the tasks presented in the traditional TSST without the stressful elements. Then, for the final test of the vocabulary words, participants typed the Hungarian word equivalents for Swahili words. Szöllősi et al. (2017) learned that regardless of stress-condition, participants who were in the testing condition performed better on the final test than

participants in the study condition. Moreover, Szöllösi et al. (2017) concluded that the addition of feedback to their study prior to the final test allowed participants to increase their word pair retention during the final test, which is consistent with the findings of Jang and Marshall (2018).

As concluded by Karpicke and Roediger (2008), restudying is less effective than retesting once learners know the translation of a word. In their study, Karpicke and Roediger (2008) had college-aged participants learn 40 Swahili-English word pairs. This experiment had four conditions, the traditional “study” and “test” conditions as well as two word conditions: the words in the list would either be dropped from the word pair list after they were first recalled or would be subject to retrieval practice. One week after the experiment began, participants returned and were given a final test. While there was not a significant difference among the conditions for the amount of words that were recalled during the learning session, during the final session there was a significant difference. According to Karpicke and Roediger (2008), the participants in the test condition and the retrieval practice condition correctly recalled approximately 80% of the word pairs during the final testing stage while in the other studying conditions, students recalled only 33% and 36% of the word pairs. This study demonstrates that although restudying and retesting both may seem to be beneficial to students, retesting is the most effective way to recall learned material over time.

This phenomenon can also be explained by van den Broek, Takashima, and Segers (2013)’s suggestion that testing “involves more effortful” and deeper semantic processing of material and involves more “cognitive effort” than restudying (van den Broek et al., p. 94). This added cognitive effort that testing requires is beneficial for long-term retention. In their



study, van de Broek et al. (2013) used fMRI while having Dutch participants either restudy or undergo testing on 50 Swahili-Dutch word pairs (restudy) or Swahili words (testing) that had previously been encoded. Participants were then asked to return after seven days and complete a final computerized memory test wherein they were asked to provide the Dutch translations for Swahili words. Participants more correctly and more quickly answered the final memory test with the Dutch translations if they had been in the group that was tested on the material than the group that had restudied the material. According to van den Broek et al. (2013), higher vocabulary retrieval after testing can be attributed to greater activity to parts of the brain which are associated with semantic processing: the middle temporal gyrus, the inferior parietal lobe, and the inferior frontal gyrus during testing. This brain activity does not occur to the same magnitude during restudying.

Although the testing effect is a phenomenon consistently validated in laboratory settings, it is not often employed in traditional foreign language learning settings, despite the educational benefits it is perceived to provide students. According to Buchin and Mulligan (2019), “the debate over the generalizability of retrieval practice to meaningful educationally valid materials remains ongoing” (Buchin & Mulligan, 2019, p. 2). For example, applying the testing effect in an educational context may also transfer the distractions that students face every day into the results of an experiment. Therefore, Buchin and Mulligan (2019) researched the validity of the testing effect under divided attention. In Experiment 1 of their study, Buchin and Mulligan (2019) exposed 40 participants to 64 Swahili-English word pairs. Participants were divided in two conditions: divided attention (DA) and full attention (FA) and within those conditions participants were assigned a restudy or retrieval condition. Participants in both DA conditions were told to pay attention

to digits (1-9), which played on their headphones once every 1500ms. Each time they heard the digit, they were instructed to use the computer keyboard to press a key to indicate whether the number was even or odd. All participants were instructed to either study or test themselves on the words in two phases: in phase 1, all 64 words were presented, and in phase 2, 16 word pairs were shown at a time, either once or thrice. Participants were asked to read the Swahili words silently but the English translations out loud. Two days later, participants were given a final test. Participants in the retrieval practice condition outperformed their peers in the restudy condition, demonstrating the typical testing effect. However, results indicate that the testing effect was similar for both DA and FA conditions, thus suggesting that the type of attention that one exhibits does not play a vital role in eliciting the testing effect.

A factor that determines a student's benefit from a testing effect may include the students' own self-regulation skillset. Ariel and Karpicke (2018) presented 60 undergraduate students with 20 Lithuanian-English word pairs. Half of the participants were in a "neutral instruction control group" condition and half were in a "retrieval practice group" condition. The researchers explained to the participants in the retrieval practice group the benefits of repeated testing (which they termed repeated retrieval practice) in an effort to induce self-regulated learning, and the participants in the "neutral instruction control group" were only told to memorize the word pairs. Participants were presented with an online program that allowed them to study the word, test themselves on the word, and finally, drop the word from their study pile once they *felt* that they had memorized the word pair. All subjects were given a final test 45 minutes after they had studied and tested themselves on the words. After the final test, researchers learned that students who were told about the benefits of repeated

self-assessment were more likely to perform repeated retrieval practice on the words, whereas the control group preferred utilizing restudying or dropping the word from their study pile to retrieval practice. Participants who were told of these benefits also outperformed the control group on the final test. This data suggests that when students are made aware of the benefits of the testing effect, they are likely to outperform their peers who were not made aware of these benefits.

In some of the aforementioned studies that tested for long-term retention, the delay between initial and final testing was up to one week (Hinze & Rapp, 2014; Roediger & Karpicke, 2006, Szöllösi et al., 2017). Long-term retention is particularly important in foreign language classrooms because a limited vocabulary or grammar greatly limits how students are able to communicate in the foreign language. In fact, foreign language learning requires a specific type of vocabulary retention. However, many prior studies consider long-term retention as occurring within a single learning session; this is insufficient because “a manipulation that improves 1-week retention from 20 to 40% would still yield failure-level performance on a course exam” (Rawson, Vaughn, Walsh, & Dunlosky, 2018, p. 58). Therefore, it is important to consider the interval between the first and last sessions of studies that focus on the testing effect and whether or not they reflect a traditional classroom-learning environment. Unfortunately, when a study is reflective of a foreign language classroom environment, an additional aspect of foreign language learning can be introduced for the participants: anxiety.

#### *State Anxiety and its Effects on Testing*

For students with test anxiety, their anxiety *is* directly related to the importance of the task if it is an educational assessment. Students with high levels of anxiety may have

adverse experiences with test-enhanced learning and testing in general (Putwain, 2008). Moreover, there is not just one type of anxiety that may be created by testing: rather, there are two main categories of anxiety, trait anxiety and state anxiety (MacIntyre & Gardner, 1994). Trait and state anxiety differ from each other in that trait anxiety is a predisposed condition that occurs in a variety of situations wherein one's anxiety level is higher than what is typical but is relatively stable for the individual (MacIntyre & Gardner, 1994). Conversely, state anxiety is a type of anxiety that occurs in *specific* situations. Test anxiety, in addition to foreign language classroom anxiety, is a type of state anxiety. According to Putwain (2008), research indicates that test anxiety differs from general anxiety because it occurs in situations that require evaluation, both due to the potential threat of failure and to the worry of lowered self-esteem.

Hinze and Rapp (2014) conducted an experiment to test the effects that state anxiety has on quiz and test performance by quizzing participants either under low-stake or high-stake conditions prior to a final test. Participants in this study were told to study selected topics from a college-level biology textbook. Participants were told after studying that they were eligible to receive a potential \$5 bonus with a catch: Participants in the low-stakes quiz group were told that they would earn the bonus regardless of their performance, whereas participants in the high-stakes quiz group were told that they would receive the bonus only if they and a pre-determined partner performed better on the quizzes than the university average. After seven days, all participants were told that they would be receiving the \$5 bonus. Subsequently, they completed the final test, which consisted of novel (but related) multiple-choice items that focused on the transfer of material retention and comprehension. While performance on the quizzes was relatively equivalent for participants in the low-

stakes and high-stakes quizzing conditions, performance on the *final test* was significantly worse for participants in the high-stakes quizzing condition, despite the assurance that all participants would receive the \$5 bonus. Participants in high-stakes stress condition performed worse than the participants in the low-stakes stress condition. Hinze and Rapp (2014) predicted a learning disruption hypothesis, which states that participants who experience pressure during quizzing sessions, regardless of their performance on the quizzes, will perform worse on the final test than they would without the added pressure. Presumably, the low-stakes quiz group learned the material without added stress, and therefore performed *better* on the final test than their high-stakes counterparts who learned the material with added stress, suggesting that stress has an inverse relationship with final test performance.

While students with high cognitive test anxiety consistently perform worse at the testing stage than their low anxiety counterparts, research suggests that repeated testing has been shown to reduce students' self reported anxiety when the tests are low-stakes, or rather, do not count for a significant part of a student's grade (Cassady, 2004; McDaniel, Agarwal, Huesler, McDermott, & Roediger, 2011). In their study, McDaniel et al. (2011) created quizzes for middle school students within a science classroom prior to learning material, after learning the lesson, and one day before their unit exam. Quizzes counted for a low proportion, 10%, of the students' final grade in class. According to McDaniel et al. (2011), unit exam scores increased from an average of 79% to greater than 90%, demonstrating the testing effect. Furthermore, 64% of students responded via self-report that the frequent quizzing had reduced their anxiety levels prior to taking their unit exam.

*Foreign Language Anxiety and its Relationship to Language Learning*

Anxiety plays a significant role during foreign language learning. Thought of as a “profoundly unsettling psychological proposition” by researcher Alexander Guiora, foreign language learning is well known to produce anxiety in nonnative speakers (Horwitz et al., 1986, p. 126). Anxiety in general is known to negatively affect performance, but Horwitz et al. (1986) set out to further understand the anxiety that is associated with foreign language learning specifically. The researchers first focused on anxieties related to foreign language learning, which include the apprehension of communicating in a foreign language, test anxiety, and the “fear of negative evaluation” (Horwitz et al., 1986, p.127). Horwitz et al. created the Foreign Language Classroom Anxiety Scale (FLCAS), which is an in-depth self-report of anxiety specific to foreign language learning. According to FLCAS, a score of 33-75 indicates low levels of foreign language classroom anxiety, a score of 76-119 indicates moderate levels, and a score of 120+ indicates that an individual has a high level of foreign language classroom anxiety. Horwitz et al. (1986) concluded that anxiety is an applicable emotion to the majority of students in beginning language classes. Moreover, Horwitz et al. (1986) suggested that if students feel less test anxiety and foreign language classroom anxiety in their foreign language classrooms they would perform better on exams.

These assertions are supported by the results found by Gardner and Gardner (1997), who recruited participants from a five-week intensive second language summer program in Canada. Participants were students between grades 9-12. Students self-reported their French language proficiency and had their oral proficiency determined via speech samples by researchers at the beginning of the study. Students also self-reported for 23 questions in evaluative questionnaires (including ones about anxiety) and assessed their French proficiency with the remaining 8 questions. The results suggest that beginner French

students were significantly more anxious than intermediate students (who were, in turn, more anxious than advanced students). Gardner and Gardner (1997) proposed that there is an inverse relationship between anxiety about speaking French and an individual's proficiency/training in French.

Foreign language anxiety may be distinctive from test anxiety, although test anxiety is a component of FLCAS. Salehi and Marefat (2014) explored whether there was any relationship between foreign language anxiety, test anxiety, and foreign language exam performance. Salehi and Marefat (2014) recruited Pre-Intermediate 1 EFL students at the Iran Language Institute, Karaj Branch, as their participants. Participants were first asked to self-report their level of anxiety using the FLCAS developed by Horwitz et al. (1986). In this questionnaire, participants were asked to report their "feelings and attitudes about their English class" (Horwitz et al., 1986, p. 934). Participants were also asked to self-report their test anxiety levels with the Test Anxiety Scale (Sarason, 1975). This questionnaire was used to determine participants' feelings about any upcoming exams and was therefore not specific to their English class. Salehi and Marefat (2014) found that there was a negative relationship between foreign language anxiety and test performance on the students' Pre 1 English final exam. The researchers found that the relationship between test anxiety and foreign language test performance was negative as well. Finally, Salehi and Marefat (2014) found evidence that there was a positive relationship between foreign language test anxiety and test anxiety; they suggest that this may be due to the fact that participants responded to the Test Anxiety Scale specifically about their experiences with foreign language testing rather than to any class in which they had a test.

Conversely, results from a study which examined the correlation between foreign language learning anxiety, test anxiety, and academic performance, suggest that foreign language classroom anxiety and test anxiety are distinctive in terms of students' language achievement. Cakici (2016) recruited participants learning English as a foreign language from the Ondokuz Mayıs University in Turkey. Cakici (2016) measured participants' test anxiety using the Test Anxiety Inventory (Spielberger, 1980), and participants' foreign language classroom anxiety was measured by using FLCAS. Then, Cakici (2016) measured participants' academic language performance by averaging their scores at the end of the year; a score of 60 or above indicated achievement in English, while a score of 59 or below indicated poor performance in English. There was a positive correlation between foreign language class anxiety and test anxiety; however, they differentially correlated with academic achievement. While there was not a significant correlation between test anxiety and academic achievement, there was a significant negative correlation between foreign language classroom anxiety and academic achievement. These results may suggest that test anxiety and foreign language classroom anxiety are distinct when it comes to student academic achievement and that foreign language classroom anxiety may be a better predictor of academic achievement than generalized test anxiety.

Learning vocabulary is integral in foreign language learning at the beginner stage of a foreign language. MacIntyre and Gardner (1994) aimed to address the role that induced anxiety plays in vocabulary retention and performance. To do so, researchers *filmed* participants as they learned and demonstrated their knowledge of English-French noun-pair associations in order to incite anxiety within the participants. Participants, who were students in first-year French courses at the university level, were told that the words they



would be learning would be used later on in the study. Participants were split into four groups, which consisted of three groups that were each exposed to a camera beginning at one specific time point: prior to “the input”, during “processing”, or at the “output stage of learning, respectively”. The control group was never exposed to the camera. The input stage required participants to view a French noun on a computer, the processing stage required the participant to view the French noun paired with an English translation, and the output stage of learning required participants to type the French noun when given the English translation. Prior to the experiment, participants were presented with scales to self-assess their anxiety levels during the input, processing, and output stages of the experiment. The results indicated that participants were most anxious immediately following the advent of the camera regardless of the stage at which the camera was introduced. Participants’ performance on typing the French equivalent of an English word decreased for participants in each of the three stages who had “most recently had [their] anxiety aroused”, although performance was worst following the advent of the camera at the input stage (MacIntyre & Gardner, 1994, p. 15). This may suggest that there is a negative relationship between the advent of a stressor that induces state anxiety and test performance. Although this study did not differentiate between general anxiety and foreign language classroom anxiety, it models the negative effects that anxiety can have on students’ learning and subsequent performance on tests.

The role that foreign language anxiety plays as the testing effect occurs is not well documented. Although anxiety can sometimes serve as necessary motivation for students to perform well, anxiety caused by either foreign language learning or by testing can be detrimental to students’ learning and future retention (Salehi & Marefat, 2014). Simply

“coping with such anxiety uses valuable mental resources that could otherwise be better used to meet the cognitive demands of the task at hand, hence leading to poor task performance” (Mok & Chan, 2015). Moreover, according to Horwitz et al. (1986), even when students receive the same grade on an exam, an anxious student will likely have studied less efficiently due to their anxiety (and subsequently spend more time studying) than one who is less anxious.

However, there are outside factors that may increase a students’ level of anxiety in a foreign language classroom. Linguistic familiarity with foreign languages may play a role in determining students’ anxiety levels (Dewaele, 2007). For example, as a language with Latin roots, English shares more linguistic similarities to French than it does with a Semitic language like Arabic. In terms of ease of learning, languages with cross-linguistic similarities aid in comprehension, production, and understanding sentence structure/grammar (Ringbom, 2007). Languages with zero relation, however, do not aid with these factors; rather, learners must take longer to learn the alphabet and will not be able to transfer the vocabulary and grammar from their L1 to their L2. However, research also suggests that previous experience with *any* foreign language can impact one’s level of anxiety about communicating in a different foreign language. Dewaele (2007) asked monolingual, bilingual, trilingual, and polyglot speakers of various languages to rate their communicative and foreign language production anxieties on a scale of one to five. Monolinguals and bilinguals experienced the highest rates of foreign language anxiety while trilingual speakers and polyglots experienced significantly less anxiety during foreign language production. Although this study did not control for the languages that people were fluent in, it suggests that multilingual students may be less anxious about foreign language

learning than their monolingual peers by virtue of their experience with other foreign languages.

The hypothesis for this experiment is that the testing effect will occur for students with lower levels of foreign language classroom anxiety but will not occur for students with higher levels of foreign language classroom anxiety. If this hypothesis is supported, educators might be motivated to modify their methods of testing to be more inclusive and beneficial to students who are prone to higher levels of anxiety in their foreign language classrooms.

## **Method**

### **Participants**

Participants recruited for this study were students enrolled in introductory foreign language classes at the Claremont Colleges, specifically Introductory French and Introductory Arabic. Both Introductory Arabic and Introductory French were offered in the Fall of 2019 at 10am and 11am, and one additional Introductory French course was offered at 1:15-2:05pm. Introductory French was also offered in the Spring of 2020 at 11-11:50am. Students were recruited specifically from these classes to participate in this study. Participants were compensated for their participation with \$10 for attending the first session and \$5 for attending each subsequent session. A maximum of \$25 was given for completing this study. Participants with prior foreign language exposure or cross-linguistic exposure were not excluded from this study. Participants initially were comprised of 33 college-age students (29 female, 4 male), but only 12 completed all four sessions and 4 completed three of the four sessions (Sessions 1, 2, and 4), resulting in 16 participants (12 female, 4 male) to

be included in the analysis. Half were Introductory French students and half were Introductory Arabic students.

### **Materials**

All participants responded to the questionnaires electronically, and each student was assigned an identification number between 1 and 33 so that they could be referenced anonymously throughout the experiment.

Prior to beginning the experiment, all students completed the following two self-report measures:

- 1) A modified Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007; see Appendix A) about the participant's language history and familiarity with their L1 and target foreign language. This questionnaire differentiates between languages that have been acquired in a home environment or learned in a classroom environment.
- 2) The modified Foreign Language Classroom Anxiety Scale questionnaire (Horwitz, 1986; see Appendix B), which assesses participants' baseline foreign language classroom anxiety levels. Participants were categorized into two anxiety groups, higher or lower anxiety, based on their responses to this questionnaire. Horwitz (1986) determined the anxiety cutoffs for FLCAS such that scores of 33-75, 76-119, and 120-165 traditionally referred to low anxiety, moderate anxiety, and high anxiety participants, respectively. The lowest and highest anxiety score in this experiment were 66 and 149 with a median score of 93. Therefore, the anxiety cutoffs for this study changed to a score below 93 signifying a lower anxiety participant and a score of 93 or above signifying a higher anxiety participant.

The vocabulary included in this study included 40 French/Arabic-English translation pairs that had not been taught prior to this experiment (see Appendix C). Moreover, the vocabulary assessment that students underwent exclusively contained words that were not cognates in other languages. This means that in the French condition, vocabulary words used in testing included word pairs like “nuage/cloud” or “papillon/butterfly” because they do not have cognates similar to Latin-based words. This was inherently a simpler feat in Arabic because fewer students had prior exposure to Semitic languages. However, Arabic vocabulary words were controlled for in terms of cognates in the same way that French words were.

### **Procedure**

The design of this experiment was a 2 (Foreign Language Classroom Anxiety: low, high) x2 (Learning Condition: SSST (study, study, study, test), STTT (study, test, test, test) between-subjects design. Participants were randomly assigned to either SSST or STTT learning conditions so that half of the participants in each anxiety group were given SSST and half were given STTT. Participants were asked to complete four sessions, each of which was spaced one week apart. Cumulatively, there were 21 days between Session 1 and Session 4.

During Session 1, all participants were exposed to the 40 word pairs one at a time via computer using a Qualtrics survey, where each pair was shown for 45 seconds for a total of 30 minutes. Participants were instructed to study each pair with the goal of the exercise to remember as many vocabulary words as possible. Sessions 2 and 3 in the SSST condition consisted of participants studying the word pairs for a maximum of 45 seconds on the same Qualtrics survey, although they had the option to press a key to move forward after viewing

a word for 35 seconds. For the STTT conditions in Sessions 2 and 3, participants were tested using a Qualtrics survey, where a French or Arabic word was presented, and they had a maximum of 30 seconds to type in the English translation for each question. Regardless of participant response, feedback was given by displaying the correct answer for 10 seconds after the 30-second maximum typing period had passed. Because participants in the STTT condition were able to press a key before the 30 seconds had passed, they were therefore exposed to each foreign language word for a *maximum* of 40 seconds. At the end of each session, regardless of condition, participants reported their anxiety levels on a scale between 0-5.

The first session lasted one hour, and the subsequent three sessions lasted for approximately half an hour. After the final session, participants were debriefed and thanked for their participation.

## Results

### *Recall of Foreign Language Vocabulary*

A univariate ANOVA with Learning Condition (SSST, STTT) and Anxiety Type (lower, higher) as between-participants variables was conducted on percent correct recall of foreign language vocabulary in Session 4, the final session. Means and standard deviations from this analysis are shown in Table 1.

Anxiety Type	Condition	<i>M</i>	<i>SD</i>	<i>N</i>
Higher	SSST	12.50	7.07	2
Lower		47.92	18.53	6
Higher	STTT	49.17	7.22	3
Lower		38.00	30.07	5

*Table 1.* Percent correct recall of foreign language word pairs in Session 4 as a function of Learning Condition and Anxiety Type

The three-way interaction, Learning Condition x Anxiety Type x Session, was not significant,  $F < 1$ , nor were the two-way interactions between Anxiety Type x Session,  $F(2, 24) = 2.43$ ,  $MSE = 1.29$ ,  $p < 0.16$ , or Learning Condition x Session,  $F(2, 24) = 1.59$ ,  $MSE = 0.84$ ,  $p < 0.24$ . The two-way interaction between Condition x Anxiety Type was marginally significant,  $F(1, 12) = 3.95$ ,  $MSE = 457.45$ ,  $p < 0.07$ . There was no main effect of Learning Condition,  $F(1, 16) = 1.30$ ,  $MSE = 457.45$ ,  $p < 0.28$ , or Anxiety Type,  $F(1, 16) = 1.07$ ,  $MSE = 457.45$ ,  $p < 0.32$ .

Following up the interaction between Learning Condition and Anxiety Type (see Figure 1), an independent samples t-test showed that for higher anxiety participants, there was a significant effect of learning condition,  $t(3) = -5.60$ ,  $SE = 6.54$ ,  $p < 0.01$ , showing a testing effect such that higher anxiety participants in the STTT condition recalled more in Session 4 than the higher anxiety participants in the SSST condition. In contrast, there was no significant effect of learning condition for lower anxiety participants,  $t < 1$ . Within learning condition, an independent samples t-test showed that for participants in the SSST learning condition, there was a significant effect of anxiety type,  $t(6) = 2.53$ ,  $SE = 14.01$ ,  $p < 0.045$ , showing that lower anxiety participants had greater recall in Session 4 than higher anxiety participants. However, there was no significant effect of anxiety type for participants in the STTT condition,  $t < 1$ , showing equivalent recall for higher and lower anxiety participants.

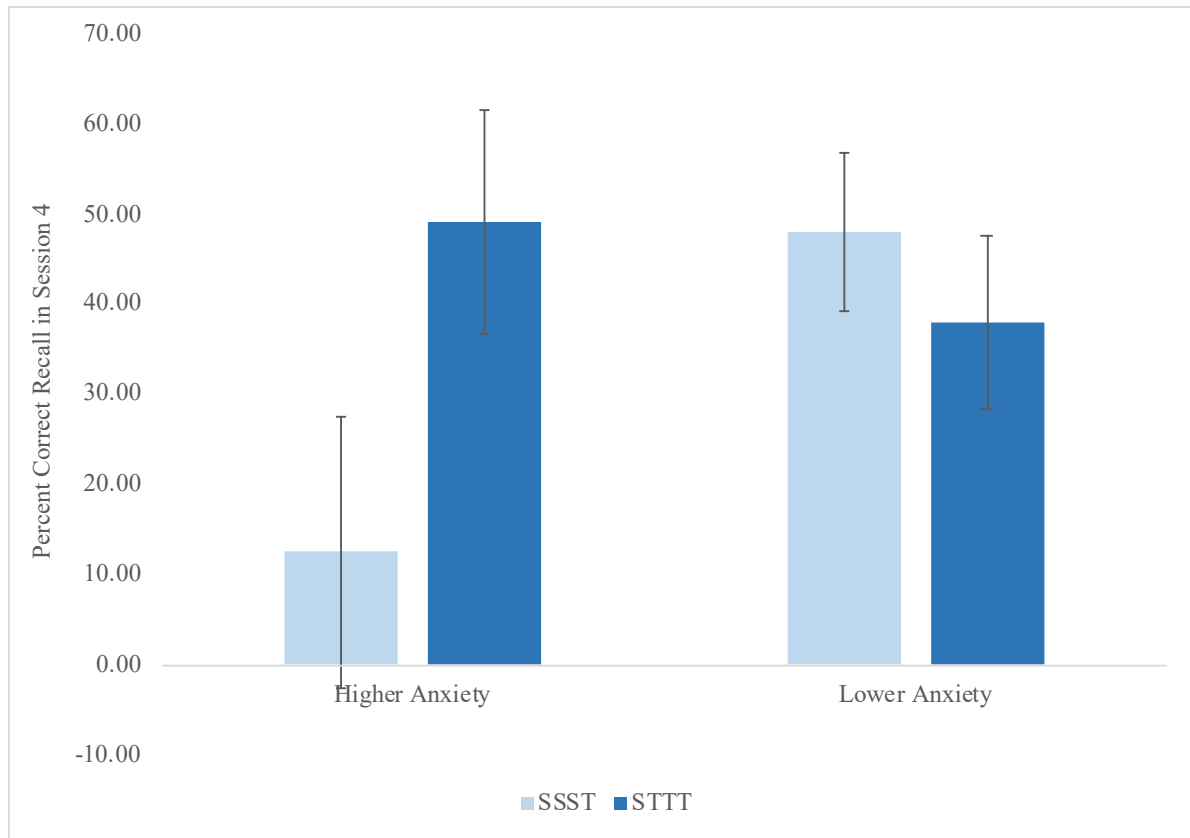


Figure 1. Average percent correct recall as a function of learning condition and anxiety type

The next analysis looked at the testing effect as a function of the two target languages studied. A univariate-measure ANOVA with Target Language (French or Arabic) and Learning Condition (SSST, STTT) as between-participants variables was conducted on percent correct recall of foreign language vocabulary in Session 4. The two-way interaction between Target Language and Learning condition was not significant,  $F < 1$ . While the main effect of Learning Condition was not significant,  $F < 1$ , the main effect of Target Language was marginally significant,  $F(1, 12) = 8.75$ ,  $MSE = 361.72$ ,  $p < 0.012$ , showing that participants whose target language was French had greater recall than participants whose target language was Arabic (see Figure 2).



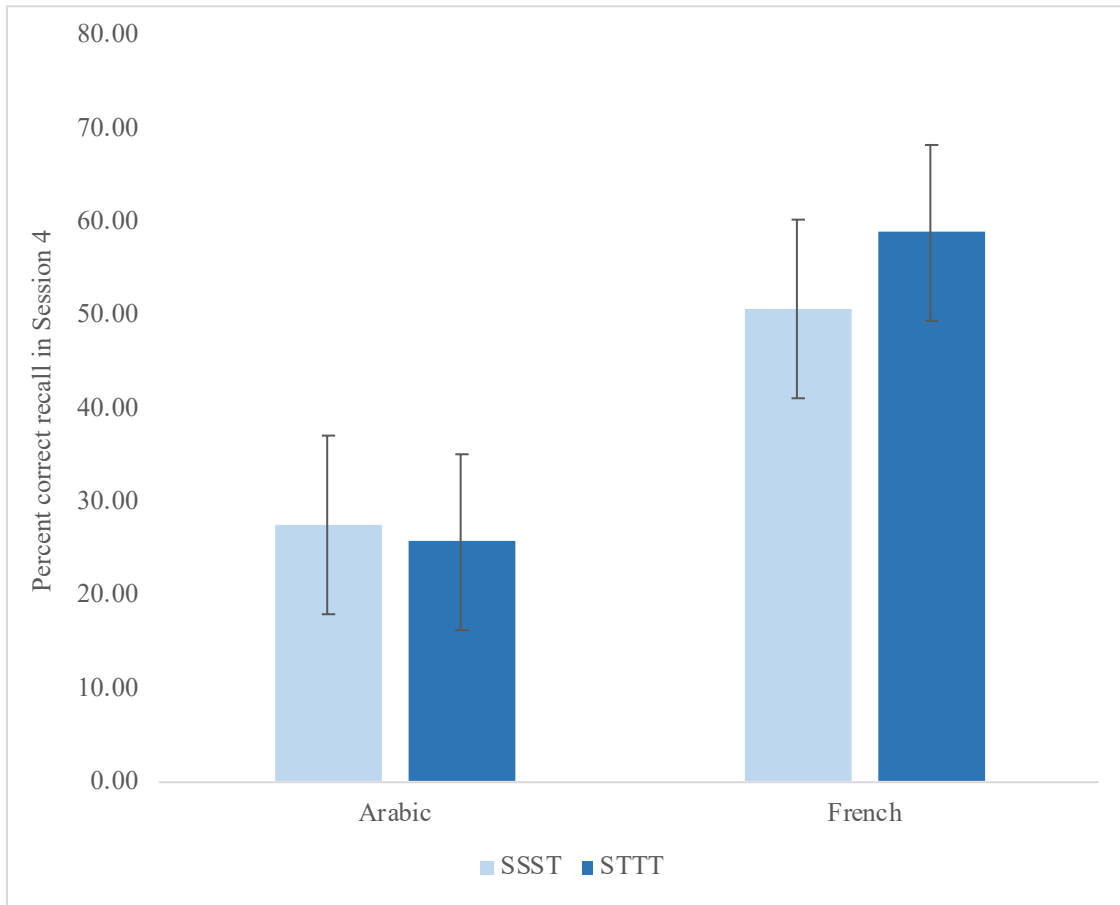


Figure 2. Percent correct recall at Session 4 as a function of Target Language and Learning Condition

### ***Correlations Between LEAP-Q Responses and Recall***

Self-reported data from the LEAP-Q, specifically the number of languages spoken and the number of cultures with which one identified, were analyzed using Pearson correlations to assess whether there were any relationships between these variables and percent correct recall in Session 4. There was a marginally significant negative correlation between number of cultures with which one identified and percent correct recall in Session 4,  $r(16) = -0.42$ ,  $p < 0.10$ , suggesting that as the number of cultures with which a participant identified increased, their percent recall decreased. The correlation between number of languages spoken and percent recall was not significant although it trended in the positive direction,  $r(16) = 0.30$ ,  $p < 0.25$

The above correlations were further explored separately for each learning condition. These analyses revealed that correlations with percent recall in Session 4 were stronger for participants in the STTT condition, who had a marginally significant positive correlation between number of languages spoken and percent recall,  $r(8) = 0.64, p < 0.09$ , and a marginally significant negative correlation between number of cultures and percent recall,  $r(8) = -0.63, p < 0.10$ . In contrast, for participants in the SSST condition, there were no significant correlations between number of languages spoken and percent recall,  $r(8) = -0.02, p < 0.97$ , or between number of cultures and percent recall,  $r(8) = -0.40, p < 0.331$ , although this negative correlation was relatively large and consistent with that observed for participants in the STTT condition.

Pearson correlations were also explored within each target language. For participants whose target language was French, there was no significant correlation, either between the number of languages spoken and their percent recall,  $r(8) = 0.08, p < 0.86$ , or between the number of cultures and percent recall,  $r(8) = -0.37, p < 0.37$ , although this was in the same negative direction as previously reported correlations of this type. For participants whose target language was Arabic, the correlation between the number of languages and their percent correct recall was marginally significant,  $r(8) = 0.62, p < 0.1$ , whereas the correlation between number of cultures and percent recall was nonsignificant,  $r(8) = -0.20, p < 0.63$ .

Additionally, participants' self-reported data for the percent of time a participant was exposed to their L1 as opposed to the other languages that they knew and their perceived proficiency in reading and speaking with their L1 were analyzed using Pearson correlations with percent correct recall in Session 4. There was a negative correlation between the

percent time a participant was exposed to their L1 as opposed to other languages and percent recall,  $r(16) = -0.61, p < 0.013$ . There was also a strong negative correlation between the time a participant spent reading in their L1 as opposed to other languages and percent recall,  $r(16) = -0.70, p < 0.003$ . The correlation between percent time speaking and percent correct recall,  $r(16) = -0.29, p < 0.27$ , was nonsignificant.

### ***Self-Reported Anxiety Levels***

A repeated-measure ANOVA with Learning Condition (SSST, STTT) as the between-participants variable and Session (1, 2, 3, 4) as a within-participants variable was conducted on mean self-rated anxiety level. Means and standard deviations from this analysis are shown in Table 2. The two-way interaction between Learning Condition and Session was not significant,  $F(3, 30) = 1.50, MSE = 0.70, p < 0.23$ , nor was the main effect of learning condition,  $F < 1$ . However, the main effect of Session was marginally significant,  $F(3, 30) = 2.25, MSE = 0.70, p < 0.10$ . Pairwise comparisons showed that anxiety levels were lower in Session 3 relative to all of the other sessions,  $ps < .05$ , with no differences between Sessions 1, 2, and 4.

Condition	Session	<i>M</i>	<i>SD</i>	<i>N</i>
SSST	1	1.33	0.72	8
SSST	2	1.50	0.42	8
SSST	3	0.67	0.41	4
SSST	4	2.00	0.68	8
STTT	1	2.00	0.72	8
STTT	2	1.67	0.42	8
STTT	3	1.17	0.41	8
STTT	4	1.33	0.68	8

*Table 2.* Means and standard deviations from analysis of anxiety levels

Although the significance of these results is limited by the small sample, there are trends within the data that are worth noting (see Figure 3). First, the mean anxiety level of

participants in both the SSST and STTT conditions generally decreased across the first three sessions. However, in Session 4, the anxiety of participants in the SSST condition increased more than the anxiety of participants in the STTT condition.

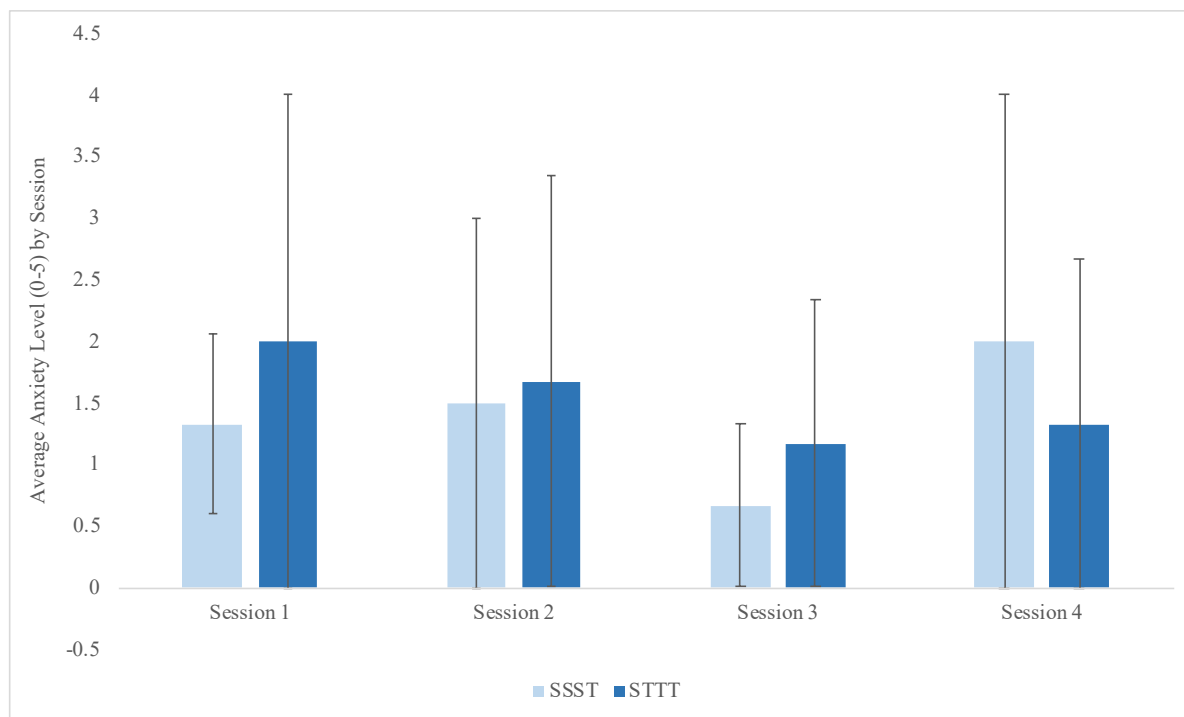


Figure 3. Average Anxiety Level (scale from 0-5) over Sessions as a Function of Learning Condition

Independent of learning condition, a mixed-measures ANOVA with Target Language (French or Arabic) as a between-participants variable and Session (1, 2, 3, 4) as a within-participants variable was conducted on mean anxiety levels, but there were no significant effects. The two-way interaction between Target Language and Session was not significant,  $F(3,30) = 1.17$ ,  $MSE = 0.72$ ,  $p < 0.34$ . The main effects of Session,  $F(3,30) = 2.05$ ,  $MSE = .72$ ,  $p < 0.13$  and Target Language,  $F < 1$ , were also nonsignificant.

Target Language	Session	<i>M</i>	<i>SD</i>	<i>N</i>
Arabic	1	1.25	0.89	8
	2	2.00	0.50	8
	3	1.00	0.52	4
	4	2.00	0.84	8
French	1	1.88	0.63	8
	2	1.38	0.35	8
	3	0.88	0.37	8
	4	1.50	0.59	8

*Table 3.* Means and standard deviations of anxiety levels in Sessions 1-4 reported from analysis of Target Language

### ***Correlations between LEAP-Q Responses and Anxiety Level***

Participants' self-reported data on the number of languages spoken and number of cultures identified with were analyzed with Pearson correlations to see if there was a relationship with anxiety level in Session 4. There was a significant negative correlation between the number of languages spoken by the participant and their anxiety level in Session 4,  $r(16) = -0.52, p < 0.039$ , suggesting that as the number of languages spoken increased, their anxiety level decreased. The relationship between the number of cultures a participant identified with and their anxiety level at Session 4 was nonsignificant,  $r(16) = 0.32, p < 0.23$ .

### **Discussion**

This study hypothesized that foreign language classroom anxiety would have a detrimental impact on the retrieval of vocabulary words of foreign language learners in situations that require repeated testing, contrary to the typical beneficial effects of testing that has been shown repeatedly, both in foreign language learning and other types of material (Hinze & Rapp, 2014; MacIntyre & Gardner, 1994; McDaniel et al., 2007; Roediger & Karpicke, 2006; Szöllösi et al., 2017; van den Broek et al., 2013). Previous

research has suggested that foreign language classroom anxiety and test anxiety are negatively correlated with test performance (Cassady, 2002; Horwitz et al., 1986). Thus, it was expected to find that lower-anxiety participants would have a larger testing effect, i.e., greater recall in the STTT condition than in the SSST condition, than higher-anxiety participants whose anxiety may prevent them from benefitting from testing. However, the hypothesis for this study was not supported. In fact, higher anxiety participants had a significant testing effect, whereas lower anxiety participants did not (and trended in the opposite direction). Moreover, the average anxiety level of participants decreased by Session 3 relative to the previous sessions, which for participants in the STTT condition meant that they reported less anxiety with each subsequent test. For Session 4, the final test session, participants in the SSST condition and not the STTT condition had an increase in anxiety. Together, the results of this study suggest that *repeated* testing not only benefits learning of foreign language vocabulary but can also offset the negative effects of anxiety on test performance that typically occurs in literature (Horwitz et al., 1986; MacIntyre & Gardner, 1994; McDaniel et al., 2011; Mok & Chan, 2015; Salehi & Marefat, 2014). This may be due to the relationship between state anxiety (anxiety type that occurs in specific situations and is not a predisposed condition) and test performance. While higher state anxiety levels have been negatively correlated with test performance in previous studies, repeated testing is suggested to *reduce* learners' state anxiety levels, which in turn leads to higher test performance (Cakici, 2016; Cassady, 2002; Hinze & Rapp, 2014; Horwitz et al., 1986). Foreign language classroom anxiety is a type of state anxiety because it occurs in situations that involve foreign language learning. Therefore, the present results generalize this relationship to foreign language classroom anxiety.

Within these results, however, there is a disparity in learning foreign language vocabulary as a function of the language being learned. Participants whose target language was French had greater recall in Session 4 than participants whose target language was Arabic, regardless of learning condition or anxiety type. As suggested by Ringbom (2007), languages that have a cross-linguistic relationship with a learners' L1 benefit the learners' comprehension and production of the target language. The first language acquired by all participants in this study (English,  $N = 15$ , Swahili,  $N = 1$ ) share Latin roots. Therefore, it is understandable that participants whose target language was French performed better on the final test; their familiarity with French as a Romance language, which uses the same alphabet as English, aided them in learning the new vocabulary pairs. In contrast, no participant already knew a language that shared cross-linguistic features with Arabic. This lack of shared prior knowledge of Semitic languages, combined with the added difficulty of knowing that the translation of a word pair did not share cross-linguistic features with a participant's other languages, may have contributed to students whose target language was Arabic having poorer recall. Moreover, Ringbom (2007) states that languages that do not have cross-linguistic similarities with a learners' L1 may take longer to learn and, as suggested by Karpicke and Roediger (2008), recall. The results of Karpicke and Roediger state that in terms of recall, repeated testing is more effective only once participants know the translation of the word pair, so it is possible that participants attempting to learn the English-Arabic pairs had not learned them as well in the first place, making it difficult to demonstrate a testing effect.

However, being multilingual had some benefit on recall. While there was not a significant correlation between number of languages and recall overall, there was a positive

relationship between number of languages and recall for participants in the STTT condition but not the SSST condition. This may suggest that repeated testing *enhances* the benefits that being multilingual has on recall, while repeated studying diminishes those benefits. This positive correlation between the number of languages that a participant spoke and their percent correct recall also emerged for participants whose target language was Arabic but not French. This contrasts with idea that if a learner knows multiple languages that are linguistically dissimilar, their ease of learning words will not increase in the way it would with languages that share cross-linguistic similarities (Ringbom, 2007). These differing results may be explained by the interference of the participants' other known languages on the recall of the target language vocabulary, an idea that will be further addressed in the following paragraph. It is also worth noting that being multilingual is linked to a reduction of anxiety when learning a new language (Dewaele, 2007). Correlational analyses in this experiment replicated this effect, as the number of languages spoken increased, participants' anxiety level decreased.

While cross-linguistic familiarity may aid students' learning of a target foreign language, two variables significantly negatively correlated with recall in Session 4, specifically the percent time a participant was exposed to their L1 as opposed to other known languages and the percent time spent reading in their L1. Ringbom (2007) suggests that when a learner's known languages do not transfer vocabulary or grammar with a target language, the benefits of being multilingual are offset. This may be the case for participants whose target language was Arabic, as Arabic does not share cross-linguistic similarities with English. However, for participants whose target language was French, the opposite effect may have occurred: although there were no French-English or (no known) French-Swahili



cognates in the vocabulary list, there were words in the target language that sounded similar to English words, (i.e., “poitrine” was often considered to be a word pair with “latrine”, and “timbre” was associated with “wood”, even though those English words never appeared in the word pair list). Perhaps the more time a student exposes themselves to their L1, especially in reading, the less capacity they have for recalling words in their target foreign language without allowing their L1 to interfere.

### **Limitations and Future Implications**

There are, of course, limitations to this study. One of the greatest may be participant attrition; all studies rely on participation, but due the limited eligibility requirements of this study and its longitudinal nature, this study was only able to collect data from a very small sample size across four between-subjects conditions ( $N = 16$ ). Following up with a larger sample will be necessary to see if the nonsignificant patterns demonstrated in this study become significant with more participants.

Another limitation of this study lies in its methodology: this experiment did not distinguish Test Anxiety from Foreign Language Classroom Anxiety when testing participants. Without testing both, this study cannot claim that student performance on word retrieval is solely a function of foreign language classroom anxiety, as research has suggested that there is a positive relationship between test anxiety and foreign language classroom anxiety. There is also evidence that test anxiety and foreign language anxiety both decrease test performance (Salehi & Marefat, 2014).

Moreover, while this study did test the relatively long-term effects of the testing effect (i.e., after three weeks), its application in an educational setting may be limited due to its inaccurate reflection of a traditional classroom environment. Though the length of time a

learner is given to memorize vocabulary words before an exam was authentic, the material presented to participants to learn was not. Vocabulary word-pairs constantly change on tests, and it is unlikely to assume that a professor would provide his or her students with the same material on every quiz leading up to the test.

Moreover, this study could also not fully replicate a classroom environment because the participants' professors were neither in the classroom during the time of the study nor did the results of the study have any impact on the participants' grade. Similarly, participants' learning environment in the experiment was not consistent throughout the study like it would be in a classroom; students enrolled in Arabic were tested in one of their own classrooms (used only for instruction on Wednesday evenings as opposed to the classroom they were taught in Monday-Thursday in the afternoon) in the evening, while participants enrolled in French were tested in an academic department library before noon. Introductory French students learned their vocabulary words in an unfamiliar and non-class-related environment, but the timing of this experiment (Fridays between 11-1pm) coincided fairly consistently with the timing of their regularly scheduled class periods (Monday-Thursday in the late mornings/afternoons). For Introductory Arabic students, they learned the vocabulary word pairs in a familiar classroom environment (although a classroom that they were never quizzed/tested in), just minutes after the mandatory Wednesday evening class finished. However, the timing of this experiment was likely less ideal for recall, even though it occurred just after an Arabic class, due to the lateness of the course. Perhaps in this way, time of day played a larger role than environment for participants' correct recall of vocabulary words.

The participants for this study may have also been self-selecting in some ways. That is to say, participants who continued throughout this study voluntarily chose to enter in a low-stakes (but still anxiety-inducing) environment every week for a month. These participants were never made aware whether or not they would be tested on the material prior to beginning each session. As previously mentioned, research suggests a positive relationship between test anxiety and foreign language classroom anxiety; it is possible that the idea of being tested, repeatedly or just once, would limit the number of students with foreign language classroom or test anxiety who were willing to participate in this experiment (Salehi & Marefat, 2014).

As observed by the researcher of this study, the majority of students enrolled in Foreign Language courses at the Claremont Colleges tend to be female, with an even greater number of women enrolled in French courses than Arabic courses. The participants in this study reflect this trend, with 12 females and 4 males participating. This disparity could have potentially skewed the results of this study to be biased towards how the testing effect impacts *women*, dependent on their anxiety levels. As suggested by the results of Cakici (2016), women experience significantly more foreign language classroom anxiety than men. In the present experiment, this disparity is reflected in that all of the male participants ( $N = 4$ ) were categorized into the lower-anxiety group; in contrast, the female participants comprised the whole higher anxiety group ( $N = 5$ ) and the majority of the lower anxiety group ( $N = 11$ ). Future research should consider the demographics of introductory language courses and recruit participants from courses in target foreign languages with a more comparable female-male ratio. Another way to increase the number of higher anxiety participants in a future study may be to simulate the participants' learning environment more

similarly to a foreign language classroom, with their target language professor and classmates present as they perform the restudying or retesting conditions.

The results of this study suggest that repeated testing remains an efficient and effective measurement for students with higher anxiety. Repeated testing may reduce participants' anxiety levels and, in turn, produce higher percent recall of higher anxiety students on final tests. However, this may only be useful for target foreign languages that share linguistic features with its learners' L1s. A potentially impactful way for foreign language professors to reduce foreign language classroom anxiety and test anxiety within their classrooms might be to adopt a testing model similar to that of this study, where students are provided with low-stakes quizzes prior to being given a final test. Students may experience a reduction of anxiety, particularly those who have higher foreign language classroom anxiety. Although this study was tested on French and Arabic students, it is possible that other college-level foreign language learners would benefit from a repeated testing model such as this one.

### References

- Ariel, R., & Karpicke, J. D. (2017). Improving self-regulated learning with a retrieval practice intervention. *Journal of Experimental Psychology: Applied*, *24*(1), 43–56. doi: 10.1037/xap0000133
- Buchin, Z. L., & Mulligan, N. W. (2019). The testing effect under divided attention: Educational application. *Journal of Experimental Psychology: Applied*, *25*(4), 1–18. doi: 10.1037/xap0000230
- Cakici, D. (2016). The correlation among EFL learners' test anxiety, foreign language anxiety and language achievement. *English Language Teaching*, *9*(8), 190–196. doi: 10.5539/elt.v9n8p190
- Cassady, J. C. (2004). The influence of cognitive test anxiety across the learning-testing cycle. *Learning and Instruction*, *14*(6), 569-592. <http://dx.doi.org/10.1016/j.learninstruc.2004.09.002>
- Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, *27*(2), 270-295. doi:10.1006/ceps.2001.1094
- Dewaele, J. (2007). The effect of multilingualism, sociobiographical, and situational factors on communicative anxiety and foreign language anxiety of mature language learners. *International Journal of Bilingualism*, *11*(4), 391-409. doi:10.1177/13670069070110040301
- Gardner, R. C., Smythe, P. C., & Brunet, G. R. (1977). Intensive second language study: Effects on attitudes, motivation and French achievement. *Language Learning*, *27*(2), 243-261. doi:10.1111/j.1467-1770.1977.tb00121.x

- Hinze, S. R., & Rapp, D. N. (2014). Retrieval (sometimes) enhances learning: Performance pressure reduces the benefits of retrieval practice. *Applied Cognitive Psychology, 28*, 597–606. <https://doi.org/10.1002/acp.3032>.
- Horwitz, E., Horwitz, M., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125-132.
- Jang, Y., & Marshall, E. (2018). The effect of type of feedback in multiple-choice testing on long-term retention. *The Journal of General Psychology, 145*(2), 107–119. doi: 10.1080/00221309.2018.1437021
- 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. doi: 10.1037/0278-7393.33.4.704
- Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966-968. doi:10.1126/science.1152408
- Macintyre, P. D., & Gardner, R. C. (1994). The Effects of Induced Anxiety on Three Stages of Cognitive Processing in Computerized Vocabulary Learning. *Studies in Second Language Acquisition, 16*(1), 1–17. doi:10.1017/S0272263100012560
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). Language experience and proficiency questionnaire. *Journal of Speech Language and Hearing Research, 50*(4), 940–967. doi: 10.1037/t01517-000
- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of

- quiz frequency and placement. *Journal of Educational Psychology*, 103(2), 399-414. doi:10.1037/a0021782
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19(4), 494–513. doi: 10.1080/09541440701326154
- Mok, W. S., & Chan, W. W. (2016). How do tests and summary writing tasks enhance long-term retention of students with different levels of test anxiety? *Instructional Science*, 44(6), 567-581. doi:10.1007/s11251-016-9393-x
- Pastötter, B., & Bäuml, K. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, 286(5), 1-5. doi:10.3389/fpsyg.2014.00286
- Putwain, D. W. (2008). Deconstructing test anxiety. *Emotional and Behavioural Difficulties*, 13(2), 141-155. doi:10.1080/13632750802027713
- Rawson, K. A., Vaughn, K. E., Walsh, M., & Dunlosky, J. (2018). Investigating and explaining the effects of successive relearning on long-term retention. *Journal of Experimental Psychology: Applied*, 24(1), 57–71. doi: 10.1037/xap0000146
- Ringbom, H. (2007). *Cross-linguistic Similarity in Foreign Language Learning*. Clevedon [England]: Multilingual Matters. Retrieved from <http://search.ebscohost.com.ccl.idm.oclc.org/login.aspx?direct=true&db=nlebk&AN=181421&site=ehost-live&scope=site>
- Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced learning. *Psychological Science*, 17(3), 249-255. doi:10.1111/j.1467-9280.2006.01693.x

- Salehi, M., & Marefat, F. (2014). The effects of foreign language anxiety and test anxiety on foreign language test performance. *Theory and Practice in Language Studies*, 4(5), 931-940. doi:10.4304/tpls.4.5.931-940
- Szöllősi, Á, Keresztes, A., Novák, B., Szászi, B., Kéri, S., & Racsmány, M. (2017). The testing effect is preserved in stressful final testing environment. *Applied Cognitive Psychology*, 31(6), 615-622. doi:10.1002/acp.3363
- Van den Broek, G., Takashima, A., & Segers, E. (2013). Neural correlates of testing effects in vocabulary learning. *Neuroimage*, 78, 94-102.  
doi:10.1016/j.neuroimage.2013.03.071



**Appendix A: Modified LEAP-Q Questionnaire**

1. ID number
3. Today's Date:
4. Age:
5. Date of Birth:
6. Male/Female/Other:
7. Please list all the languages you know in order of dominance:
8. Please list all the languages you know in order of acquisition (your native language first):
9. Please list what percentage of the time you are currently and on average exposed to each language (your percentages should add up to 100%):
11. When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you (your percentages should add up to 100%):
12. When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of the time would you choose to speak each language? Please report percent of total time (your percentage should add up to 100%)
13. Please name the cultures with which you identify. On a scale from zero to ten, please rate the extent to which you identify with each culture (examples of possible cultures include US-American, Chinese, Jewish-Orthodox, etc.):

French (this was the same for Arabic, but with "Arabic" replacing "French")

14. Age when you began acquiring French:
15. Age when you began reading in French:
16. List the number of years and months you spent in a country where French is spoken
17. List the number of years and months you spent in a family where French is spoken
18. List the number of years and months you spent in a school and/or working environment where French is spoken
19. On a scale from zero to ten please select your level of proficiency in speaking, understanding, and reading French:
20. Please rate to what extent you are currently exposed to French in the following contexts: Interacting with friends, interacting with family, watching TV, listening to radio/music, reading, language-lab/self instruction
21. In your perception, how much of a foreign accent do you have in French? 1-10
22. Please rate how frequently others identify you as a non-native speaker based on your accent in French. 1-10

**Appendix B: Foreign Language Classroom Anxiety Scale****Foreign Language Classroom Anxiety Scale: Created by Horwitz et al. (1986)**

**Please answer the following questions by providing the number correspondent to the option that best describe your opinion.**

- 1. Strongly Agree    2. Agree    3. Neither Agree nor Disagree    4. Disagree**  
**5. Strongly Disagree**

1. I never feel quite sure of myself when I am speaking in my foreign language class.
2. I don't worry about making mistakes in language class.
3. I tremble when I know that I'm going to be called on in language class.
4. It frightens me when I don't understand what the teacher is saying in the foreign language.
5. It wouldn't bother me at all to take more foreign language classes.
6. During language class, I find myself thinking about things that have nothing to do with the course.
7. I keep thinking that the other students are better at languages than I am.
8. I am usually at ease during tests in my language class.
9. I start to panic when I have to speak without preparation in language class.
10. I worry about the consequences of failing my foreign language class.
11. I don't understand why some people get so upset over foreign language classes.
12. In language class, I can get so nervous I forget things I know.
13. It embarrasses me to volunteer answers in my language class.
14. I would not be nervous speaking the foreign language with native speakers.
15. I get upset when I don't understand what the teacher is correcting.
16. Even if I am well prepared for language class, I feel anxious about it.
17. I often feel like not going to my language class.
18. I feel confident when I speak in foreign language class.
19. I am afraid that my language teacher is ready to correct every mistake I make.
20. I can feel my heart pounding when I'm going to be called on in language class.
21. The more I study for a language test, the more confused I get.
22. I don't feel pressure to prepare very well for language class.

23. I always feel that the other students speak the foreign language better than I do.
24. I feel very self-conscious about speaking the foreign language in front of other students.
25. Language class moves so quickly I worry about getting left behind.
26. I feel more tense and nervous in my language class than in my other classes.
27. I get nervous and confused when I am speaking in my language class.
28. When I'm on my way to language class, I feel very sure and relaxed.
29. I get nervous when I don't understand every word the language teacher says.
30. I feel overwhelmed by the number of rules you have to learn to speak a foreign language.
31. I am afraid that the other students will laugh at me when I speak the foreign language.
32. I would probably feel comfortable around native speakers of the foreign language.
33. I get nervous when the language teacher asks questions which I haven't prepared in advance.

**Appendix C: List of Foreign Language Vocabulary Words**

<b>English</b>	<b>French</b>
bean	haricot
bike	vélo
bridge	pont
butterfly	papillon
cliff	falaise
closet	poitrine
cloud	nuage
donkey	âne
drawer	drawer
factory	usine
fish	poisson
flu	grippe
food	aliment
gift	cadeau
hallway	couloir
heritage	patrimoine
hood	capot
keyboard	clavier
leather	cuir
lobster	homard
nail	clou
network	réseau
oil	huile
plane	avion
poster	affiche
raspberry	framboise
river	fleuve
roof	toit
skin	peau
sock	chaussette
stage	étape
stamp	timbre
station	gare
stove	fourneau
stripe	rayure
suit	tailleur
tablecloth	nappe
waste	gaspillage
whim	caprice
windowpane	carreau

<b>English</b>	<b>Arabic</b>
army	جيش
art	فن
award	جائزة
boat	مركب
car	سيارة
color	لون
daughter	ابنة
dinner	عشاء
divorce	طلاق
doctor	طبيب
dress	فستان
floor	طابق
fog	ضباب
gap	فراغ
garden	حديقة
gate	بوابات
irrigation	ري
knight	فارس
location	موقع
mankind	بشر
milk	حليب
mountain	هضبة
olive	زيتون
poetry	شعر
rain	مطر
religion	دين
revolution	ثورة
sand	رمال
scent	رائحة
slogan	شعار
sparrow	عصفور
theater	مسرح
thunder	رعد
tree	شجر
vein	عرق
violin	كمان
weather	طقس
window	شباك
witness	شاهد
wood	خشب