COMPUTER-ASSISTED FOREIGN-LANGUAGE VOCABULARY INSTRUCTION

by

Carson Jeffrey Smith

Submitted in partial fulfillment of the

requirements for Departmental Honors in

the Department of Psychology

Texas Christian University

Fort Worth, Texas

May 3, 2021

COMPUTER-ASSISTED FOREIGN-LANGUAGE VOCABULARY INSTRUCTION

Project Approved:

Supervising Professor: Anna I. Petursdottir, Ph.D.

Department of Psychology

Tracy Centanni, Ph.D.

Department of Psychology

Eric Simanek, Ph.D.

Department of Chemistry

ABSTRACT

A foreign language is defined as nonnative and outside of the natural linguistic community. Benefits of learning a foreign language are that it heightens employability, it positively affects cognitive functioning, and it increases cross-cultural awareness. Research on optimal presentation formats in computer-assisted instruction has been minimally extended to foreign language vocabulary learning. The goal of this study was to compare the effects of a pair-test (PT) procedure and a high-density response construction (HDRC) procedure on foreign-language vocabulary acquisition. We hypothesized that participants would learn faster in the HDRC condition and perform better on transfer and retention tests. However, no differences were found between conditions, suggesting that a greater density of constructed responses does not necessarily translate into better outcomes.

Computer-Assisted Foreign-Language Vocabulary Instruction

Learning a foreign language refers to learning a language that is nonnative and outside of the natural linguistic community (Gass & Selinker, 2001). This differs from a second language because a second language is acquired in an environment where the language is widely spoken. As a result, a foreign language is most often learned in a classroom setting or through self-study techniques, as opposed to being naturally emersed in an environment where the language is prevalent. Benefits of learning a foreign language are prevalent. The United States' demand for workers that can effectively communicate in more than one language more than doubled between 2010 and 2015 (New American Economy, 2017). On top of heightened employability, learning a foreign language positively affects cognitive functioning and increases cross-cultural awareness (Fox et al., 2019).

Students learning a foreign language need to master a vocabulary of about 3,000 frequent words to be successful in everyday situations (Nation & Waring, 1997) and 5,000 words to read text that is not simplified (Hirsh & Nation, 1992). Computer-assisted instruction can facilitate vocabulary learning (e.g., Mirzaei et al., 2015) and many computerized self-instruction programs are commercially available (e.g., Duolingo®, Rosetta Stone®, Babbel®). However, research on optimal presentation formats in computer-assisted instruction has been minimally extended to foreign language vocabulary learning.

Among aspects of vocabulary presentation that vary widely across commercially available programs are the degree of active student responding and the density of constructed-response trials. Active student responding is when a student responds to each piece of information that is presented in a way that communicates the student's learning of the material. A constructed-response trial is when the active student response consists of typing or writing out

a response, as opposed to selecting it from multiple choices. Research on using computer-assisted instruction for teaching college-level course content suggests both of these are important variables. The greater the densities of constructed-response requirements, the better students perform (e.g., Kritch et al., 1995; Kritch & Bostow, 1998) and students prefer more active student responding as opposed to less active student responding (Haggas & Hantula, 2002). However, most of these studies have evaluated instruction of complex material like definitions of concepts and terms and then their application to novel scenarios. This differs from foreign language vocabulary learning, which focuses on creating and strengthening relatively simple word-word and word-object associations. As a result, it is unknown if existing findings also hold true for teaching foreign language vocabulary.

Studies with children suggest that when learning foreign-language vocabulary, simply exposing the learner to new vocabulary words with intermittent testing (pair-test procedure) may work just as well or better than requiring an active response in each trial. In one study (Cao & Greer, 2019), children were exposed to Chinese words paired with pictures of their referents without any requirement for active responding. After learning the words through pairing, the children were able to vocalize the words in context, although only after they were first taught to echo (i.e., vocally imitate) a variety of Chinese words representing common speech sounds. In another pair of studies, Rosales et al. (2011, 2013) taught Spanish-speaking children English-language vocabulary through one-on-one instruction that was not computer-assisted. In the first study (Rosales et al., 2011), the participants initially received match-to-sample (receptive) instruction, in which an English word (e.g., the word “car”) was presented as a sample stimulus in each trial and the participants’ task was to select a corresponding picture from several options; meaning active response was required in each trial. Matching-to-sample trials like these are often

included in computerized self-instruction programs (e.g., Duolingo®). In the second study (Rosales et al., 2013), the participants were taught via modeling, in which they were simply shown pictures paired with the English words, similar to some software applications designed to expose children to foreign-language vocabulary (e.g., LittlePim®). The number of pairing trials were yoked to the number of matching-to-sample trials presented to participants in the previous study (Rosales et al., 2011). On tests for vocal naming, participants in the second study performed better (Rosales et al., 2013) than the participants in the first study.

Similar results were obtained in a study with adults (Leader & Barnes-Holmes 2001) in which the participants learned symbolic relationships between nonsense syllables (analogous to learning relationships between native- and foreign-language words) either through a match-to-sample or a pair-test procedure. Across several experiments, participants performed better at test after learning through the pair-test procedure rather than the match-to-sample procedure. This was the case both when they were tested on word-word relations that were taught directly, and on other relations that could be inferred from those taught.

Collectively, these results might suggest that when teaching foreign language vocabulary or other simple stimulus relations, a pair-test procedure may be preferable to a match-to-sample procedure that involves an active student response in each trial. However, although the match-to-sample procedures in these studies required active student responses, they did not involve constructed responses. A constructed response differs from responses in a match-to-sample procedure in that participants are required to recall the correct response in each trial, rather than just recognize it from among a number of options. A pair-test procedure also involves a requirement to recall the correct response, but only in intermittent test trials. It is possible that,

unlike matching-to-sample procedures, the more frequent recall requirement in a constructed-response procedure produces faster learning or better retention than a pair-test procedure.

Therefore, the current study sought to compare the effects of a pair-test procedure and a high-density response construction (HDRC) procedure on foreign-language vocabulary acquisition. Performance measures included speed of acquisition, transfer of learning from picture-word relations to word-word relations, and performance on a one-week retention test. Based on previous research on the density of constructed responses in computer-assisted instruction, we hypothesized that participants would learn faster in the HDRC condition and perform better on transfer and retention tests.

Method

Participants

The participants were ten undergraduate students (7 female; age range between 19 and 57 years) recruited from the TCU psychology department's human subjects pool via Sona Systems. All participants had English as their native language. Demographic information for each participant is shows in Table 1. In exchange for full participation, students were given Sona credit for their psychology courses. Participants were granted one hour of credit for completing the first part of the study. One week later, participants were given half an hour of credit for completing the retention test (A total of one and a half hours of credit). As per our IRB protocol, all participants received the gift card if they completed the follow up session. Participants who withdrew from the study prior to finishing the session were excluded from the data. Further, participants' data were also excluded if they reported that they wrote down information during the session. To be allowed to participate in this study, participants had to be at least 18 years of

age, have no previous knowledge with the language of Arabic, and have access to a quiet place with a suitable computer.

Table 1
Participant Demographic Information

Participants	Age	Gender	Ethnicity	Foreign-Language Experience (in semesters)
P1G1	19	Female	White, Non-Hispanic	Spanish (3)
P2G1	NA	Female	White, Non-Hispanic	Spanish (6)
P3G1	21	Female	White, Non-Hispanic	Spanish (6)
P4G1	20	Female	White, Non-Hispanic	Spanish (3)
P5G1	49	Female	White, Non-Hispanic	NA
P1G2	22	Male	White, Non-Hispanic	Spanish (12)
P2G2	19	Female	White, Non-Hispanic	Spanish (14)
P3G2	21	Male	Hispanic	Spanish (3)
P4G2	19	Female	White, Non-Hispanic	Spanish (4)
P5G2	57	Male	White, Non-Hispanic	Spanish (2)

Sessions were held remotely through Zoom. Participants completed the study on their computer using the remote-control feature on Zoom. They used this feature to access the SuperLab 6 software on the experimenter's computer. The experimenter was on the Zoom call so the participant could ask questions if it was needed.

Stimulus and Apparatus

Visual stimuli were presented over Zoom through the SuperLab 6 software. The timing of exposure was controlled by SuperLab as well as the responses and response feedback. Once the

experiment in SuperLab was started, the participant had remote control access to click through the study on their own accord for all of the trials.

The visual stimuli consisted of ten color photos split into five categories, with two pictures in each category. Each picture had the corresponding word in Arabic to it as well (see Appendix A). The five categories of stimuli included fruit, animals, kitchen items, food (meat), and drinks. Arabic words were selected under the conditions that they have no similar sound to the most common English word for the same item. They also were selected to have between two and three syllables. The stimuli were divided into five pairs that each included two stimuli from each category (fruit, animals, etc.). One stimulus in each pair was then randomly assigned to Set A and one to Set B.

Qualtrics was used to present the screening questions, consent document, and the follow-up test.

Procedure

Experimental Design. The study used a within-subjects pretest-posttest design combined with a single-subject multielement design. All participants received instruction of all ten words. Five words were taught via HDRC trials and five were taught via PT trials. The assignment of Set A words and Set B words was counterbalanced across participants. Before the main session began, the participant was required to fill out a form of consent on Qualtrics. The main experimental session was broken down into an instruction phase, a comprehensive pre-test, a learning phase, a comprehensive post-test, and a post-experiment questionnaire. In the instructional phase, each training sub-phase presented all ten words four times and each label test subphase presented each word one time. The comprehensive pre and post-tests were 30 trials

each. One week after the initial experiment session, the participant was emailed a link to a follow-up test on Qualtrics.

Instructions. After the participant completed the consent process on Qualtrics, the experimenter emailed the participant a zoom link where they accessed the SuperLab file via remote control access. The *Instruction Phase* occurred at the beginning of the study and the participant received the following information: "In this study, you will learn some words in Arabic. As you work through the program, try your best to remember the information given to you. Please DO NOT write down any information from the screen or take pictures of the screen to help you. Writing down information will not help you finish the experiment faster." This message stayed on the screen for 10 seconds and then the message "Press the space bar to continue" appeared. After pressing the space bar, the following two questions appeared: 1. In this study, you will be learning some words in…? (a) Portuguese (b) French (c) Arabic. 2. What were you asked NOT to do during the experiment? (a) eat or drink (b) write down information off the screen. The correct response to question 1 was "Arabic" and the correct response to question 2 was "write down information off the screen." If the participant answered either question incorrectly, they were returned to the instruction screen, then re-answered the two questions. Once they answered the two questions correctly, they read the message "You will now begin your pre-test. Press the space bar to continue."

Comprehensive Pre- and Post-Test. The comprehensive pre- and post-test served to assess performance on the word-object relations taught in the study, as well as transfer to word-word relations. 30 trials total were presented in random order. There were three trials for each word: (a) A labeling trial (b) A foreign-native translation trial (FN intraverbal) (c) A native-foreign translation trial (NF intraverbal). The *Labeling Trials* consisted of a picture that was

presented in the center of the screen. Above the picture was the instruction "What is this in Arabic?" Below the picture were instructions to "Type the word or type IDK for I don't know". These trials targeted the taught word-objects relations, whereas FN and NF intraverbal trials assessed transfer. The *FN Intraverbal Trials* consisted of a question in the form of "What is [foreign word] in English?" that appeared on the screen. Below the picture were instructions to "Type the word or select I don't know" and the participant had the option to either type into a text box or type IDK. The *NF Intraverbal Trials* consisted of a question in the form of "What is [English word] in Arabic?" that appeared on the screen. Below the picture were instructions to "Type the word or select I don't know" and the participant had the option to either type into a text box or type IDK. The participants had the option to take as long as they wanted to respond. All responses resulted in a two second intertrial interval (ITI) during which the screen was blank. There was no feedback for correct or incorrect responses during these tests.

Learning Phase. The learning phase alternated between training subphases and label test subphases, starting with a training phase. The participant was not notified when moving between subphases, trials were just presented continuously. At the beginning of the phase and after last pre-test trial, the participant read these instructions: "You will now start learning the Arabic words. Look at the pictures, read the words, and follow all instructions on the screen." A training subphase consisted of 40 trials, which was four trial blocks. Each trial blocks presented five constructed-response trials (for the five HDRC words), and five pairing trials (for the five pair-test words), in random order. When the four trial blocks were over, there was one block of label test trials and then the participant went to the next training subphase and so on.

Training Subphase: HDRC Trials. In the first trial block of the first subphase, each HDRC trial showed a picture on the screen with the foreign word directly above it. Below the picture

was a text box with the instruction “Type the word you see, then press ENTER.” After a correct response, the picture disappeared, and the word CORRECT in green letters appeared on the screen. In the second and third block of the first subphase, each HDRC trial showed a picture of the word with the first two letters of the foreign word above it and blank spaces for the rest of the letter, for example, the word “eanab” was shown as ea_ _ _. Below the picture was a text box with the instructions “Type the word, then press ENTER.” After the participant pressed enter, the picture disappeared. If the participant’s response was correct, the word CORRECT appeared in green. If it was incorrect, the word INCORRECT appeared in red and below it a statement that read, “The correct word was [correct word].” In the fourth block of the first subphase, each HDRC trial was shown the picture only, with the text box and instructions “Type the word, then press ENTER,” and the response feedback of correct and incorrect responses were exactly the same as in the second and third block. In subphases two through five, all HDRC trials were exactly like the fourth block of the first phase. Throughout all trials, after the trial is over, and after feedback there was a two second ITI with a blank screen. Then the next trial, either HDRC or pair-test, was randomly presented.

Training Subphase: Pairing Trials. Pairing trials were identical in all blocks of all subphases. A picture was presented in the middle of the screen and a word above it. The participants exited the screen by pressing a “continue” button, which went to a two second ITI with a blank screen. Then the next trial, either HDRC or pair-test was randomly presented.

Label Test Subphase. Label test trials were identical for both the HDRC and the PT condition. A picture was presented in the middle of the screen and a word above it. Below the picture was a text box with the instructions “Type the word, then press ENTER.” After the participant pressed enter, the picture disappeared. If the participant’s response was correct the

word CORRECT appeared in green. If it was incorrect, the word INCORRECT appeared in red, but the participant was NOT told the correct response. These trials were exactly like HDRC training trials except that the participant was not told the correct response after incorrect responses. Participants were considered to have reached mastery in a particular condition if they produced the words from that condition with 100% accuracy on two label tests in a row.

Post-Experiment Questionnaire. After the post-test, the participants were presented with a questionnaire that includes the following:

1. Participants rate their interest in their task and their engagement with the learning task
2. Strategies participants used to help them remember the words they were learning
3. If at any time during the experiment, they wrote down information off the screen or captured it in some other way to help them remember
4. "In this study, there were some words you were asked to type every time you saw them and other words you only had to type some of the times you saw them. If you were to re-do this study with new words and you could pick one option, would you prefer to type each word every time it is shown to you, or only some of the time?"
5. Questions about foreign-language experience, such as (a) native language, (b) do they consider themselves bilingual and (c) have they learned a foreign language in a formal educational setting

Retention Test. One week after the zoom session, the experimenters emailed the participant with a link to a Qualtrics survey containing the retention test and a code number to enter into the Qualtrics survey for the purpose of linking the retention test results to the SuperLab session data. The retention test had the same questions as the comprehensive pre-and post-test but differed in that it was 60 trials not 30 and it was not in random order. First, all labeling trials were presented

to test retention of the task that was taught directly, then all NF trials, then all FN trials. The FN trials were presented last in order to avoid exposing the participants to the Arabic words prior to completing labeling and FN trials.

Results

Figure 1 shows individual performance on the label tests during the learning phase. Filled circles represent the HDRC condition and open circles represent the pair-test condition. Three participants (P1G2, P2G2, P5G1) reached the mastery criterion in all conditions, and one (P4G2) reached criterion only in the HDRC condition. Although the remaining participants did not reach mastery, their correct responses increased across label tests. Seven participants (P1G1, P1G2, P2G2, P3G2, P4G1, P5G2, P5G1) had very similar or substantially overlapping acquisition curves in both conditions. One participant (P4G2) consistently made more correct responses in the HDRC condition and two participants (P2G1 and P3G1) consistently made more correct responses in the pair-test condition. On a paired-samples *t*-test, there was no difference between the two conditions in terms of the mean number of training phases completed before mastery was reached (the total number of training phases was used for participants who did not reach mastery), $t(9) = 1.000$, $p = 343$.

Figure 1

Label Test Performance of Individual Participants

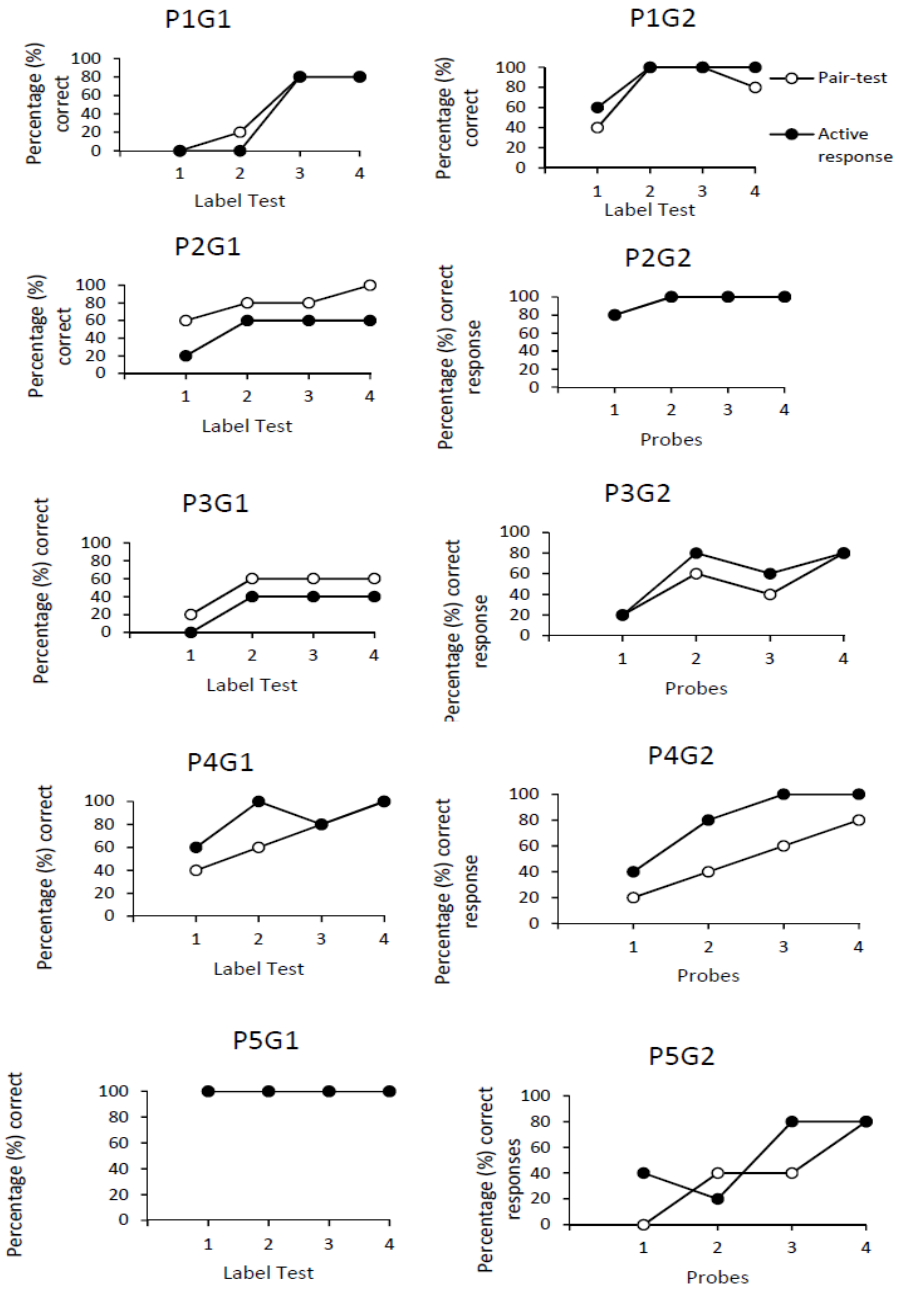


Figure 2 shows mean post-test performance. Gray bars represent the pair-test condition, while black bars represent the high-density response construction condition. On paired-samples *t*-tests, there was no significant difference between conditions on any of the three types of trials (*p*s ≥.36). Individual post-test data are shown in Appendix B.

Figure 2

Mean Scores on Post-Test

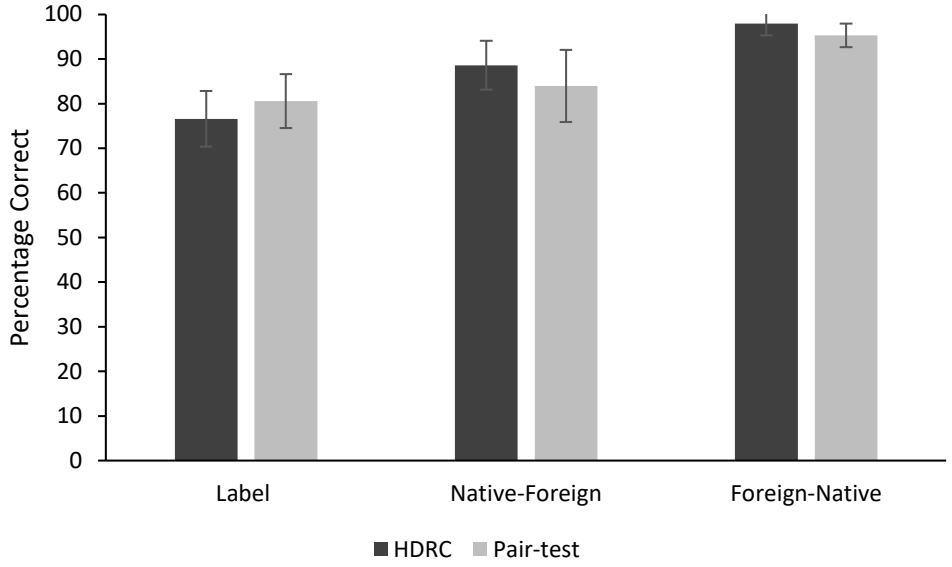
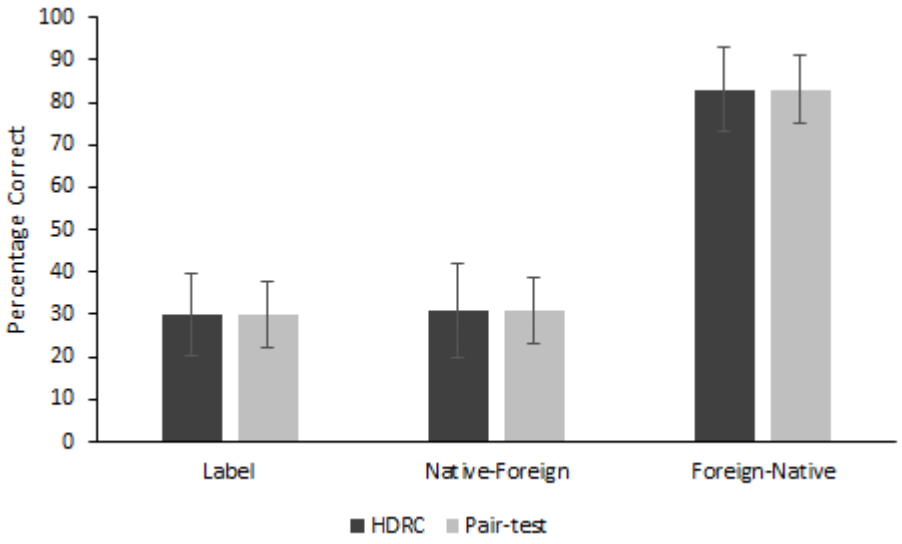


Figure 3 shows the 1-week retention-test performance for the ten participants who have completed the 1-week retention test. Gray bars represented the pair-test condition, while black bars represented the high-density response construction condition. In the retention tests, the participants were unable to recall many of the Arabic words in the label and native-foreign trials. However, once they saw the words in the foreign-native trials, they were often able to come up with the corresponding English words. Mean performance on all trial types was exactly equal in both conditions.

Figure 3

Mean Scores on One-Week Retention Test



On the post-experiment questionnaire, no participants reported writing down any information off of the screen to assist their learning. Participants reported a variety of strategies for learning the stimuli that included making associations between the words and pictures, making combinations of the first letters of the Arabic and English word, making up comedic phrases to relate the two words (Ex: Halib/Milk… "He leaped for milk") and staring at the words until they got a better grasp of the stimuli relations. Table 2 shows participants rating of their interest, engagement, and stress level, as well as which type of instruction they preferred. Interest and engagement were scored on a scale of 1-4 with 1 being not interested/engaged, 2 being slightly interested/engaged, 3 being very interested/engaged, and 4 being extremely interested/engaged. Stress level was scored on a scale of 1-4 with 1 being not stressed and 4 being extremely stressed. Overall, participants reported low interest in the task, moderate engagement, and low stress. Instruction preference was graded on a scale of 1-7, with 1 being a

preference for typing sometimes and 7 being a preference of typing all the time. Five participants reported preferring the pair-test to the HDRC condition, three preferred the HDRC condition, and two reported no preference.

Table 2

Individual Participant Reports of Experiences During Learning Task

Participants	**Interest**	**Engagement**	**Stress**	**Type of Instruction Preferred**
P1G1	2	4	2	5 (HDRC preferred)
P2G1	2	3	1	4 (no preference)
P3G1	2	3	1	7 (HDRC preferred)
P4G1	2	2	1	2 (pair-test preferred)
P5G1	4	4	1	1 (pair-test preferred)
P1G2	1	4	2	1 (pair-test preferred)
P2G2	2	3	2	2 (pair-test preferred)
P3G2	2	2	2	4 (no preference)
P4G2	2	3	2	1 (pair-test preferred)
P5G2	2	3	1	6 (HDRC preferred)
Mean	**2.1**	**3.1**	**1.5**	**3.3**

Discussion

In summary, it was found that there was no advantage of high-density response instruction relative to pair-test instruction when teaching foreign-language vocabulary. These results stand in contrast to previous research that suggested that the greater the density of constructed-response trials, the more students learned. According to past research on computer-based programmed instruction, the greater the densities of constructed-response requirements, the better students perform (e.g., Kritch et al., 1995; Kritch & Bostow, 1998). Also, students in prior

research preferred more active student responding as opposed to less active student responding (Haggas & Hantula, 2002), whereas in the present study more participants reported preferring the condition with less active student responding. However, most of the previous studies evaluated instruction of complex material like definitions of concepts and terms and then their application to novel scenarios. This differs from foreign language vocabulary learning like in this study, which focused on creating and strengthening relatively simple word-word and word-object associations. Due to this difference, the results of this study suggest that constructed-responding does not necessarily lead to better performance or learning acquisition on computer-assisted foreign language programs, at least when focusing on basic vocabulary.

Across several experiments with adults, participants learning simple stimulus associations performed better at test after learning through the pair-test procedure rather than the match-to-sample procedure (Leader & Barnes-Holmes 2001). Similar results were found in studies with children on simple word-object associations (Rosales et al. 2011, 2013). However, although the match-to-sample procedures in these studies required active student responses, they did not involve constructed responses. This study did require constructed responses and failed to show that the pair-test procedure or the constructed-response procedure had an advantage. Due to a lack of differences in the results, it suggests that a computer-assisted foreign-language vocabulary instruction program does not necessarily require either of these teaching methods over the other for the benefit of the participants who are trying to learn a foreign language.

What is interesting is the fact that participants seemed to be able to produce the correct responses in the foreign-native test on the retention test. This shows that even after a week of not actively learning the stimuli, participants were able to produce the correct English word when given the foreign word that corresponded with it. While results did not differ between conditions,

it shows that the participants still retained connections between the foreign word and the native (English) word.

Pair-test learning is less time consuming than typing a response in each trial (HDRC condition), so practically speaking, the results suggest that the pair-test procedure may be a feasible way to conduct vocabulary instruction if time is of essence, and potentially more preferred as well. Additional research is needed to demonstrate if the present results generalize to learning larger vocabularies or more complex aspects of a foreign language, such as sentence construction.

References

Alizadeh, I. (2016). Vocabulary teaching techniques; A review of common practices. *International Journal of Research in English Education, 1*(1), 22-30.

Andrä, C., Mathias, B., Schwager, A., Macedonia, M., & von Kriegstein, K. (2020). Learning foreign language vocabulary with gestures and pictures enhances vocabulary memory for several months post-learning in eight-year-old school children. *Educational Psychology Review, 32*(3), 815-850.

Cao, Y., & Greer, R. D. (2018). Mastery of echoics in Chinese establishes bidirectional naming in Chinese for preschoolers with naming in English. *The Analysis of Verbal Behavior, 34*(1), 79-99.

Fox, R., Corretjer, O., & Webb, K. (2019). Benefits of foreign language learning and bilingualism: An analysis of published empirical research 2012–2019. *Foreign Language Annals, 52*(4), 699-726.

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course* (2nd ed.). Lawrence Erlbaum.

Haggas, A. M., & Hantula, D. A. (2002). Think or click? Student preference for overt vs. covert responding in web-based instruction. *Computers in Human Behavior, 18,* 165-172.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified text for pleasure? *Reading in a Foreign Language, 8*(2), 689-696.

Kritch, K. M., & Bostow, D. E. (1998). Degree of constructed-response interaction in computer-based programmed instruction. *Journal of Applied Behavior Analysis, 31*, 387-398l

Kritch, K. M., Bostow, D. E., & Dedrick, R. F. (1995). Level of interactivity of videodisc instruction on college students' recall of AIDS information. *Journal of Applied Behavior Analysis, 28,* 85-86.

Leader, G., & Barnes-Holmes, D. (2001). Matching-to-sample and respondent-type training as methods for producing equivalence relations: Isolating the critical variable. *The Psychological Record, 51*(3), 429-444.

Mirzaei, A., Domakani, M. R., & Rahimi, S. (2015). Computerized lexis-based instruction in EFL classrooms: Using multi-purpose LexisBOARD to teach L2 vocabulary. *ReCALL, 28*(1), 22-43.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In Schmitt, N. and M. McCarthy (Eds.): *Vocabulary: Description, Acquisition and Pedagogy* (pp. 6-19). Cambridge University Press.

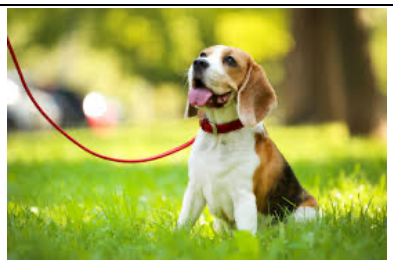
New American Economy (2017). *Not lost in translation: The growing importance of foreign language skills in the U.S. job market*. New American Economy.

Rosales, R., Rehfeldt, R. A., & Lovett, S. (2011). Effects of multiple exemplar training on the emergence of derived relations in preschool children learning a second language. *The Analysis of Verbal Behavior, 27*(1), 61-74.

Rosales, R., Rehfeldt, R. A., & Huffman, N. (2012). Examining the utility of the stimulus pairing observation procedure with preschool children learning a second language. *Journal of Applied Behavior Analysis, 45*(1), 173-177.

Appendix A

Stimuli used in the experiment. One stimulus in each group was randomly assigned to Set A and one to Set B. The assignment of Set A and Set B to experimental conditions was counterbalanced across participants.

Group:	**Category:**	**Stimulus:**	**English Word:**	**Arabic Word:**
1	Fruit		Melon	Shamam
			Grape	Eanab
2	Animals		Dog	Alkalb

			Bird	Tayir
3	Kitchen Items	 	Pan Knife	Miqla Sakin
4	Food (Meat)	 	Chicken Shrimp	Dijaj Jambiri

5	Drinks	 	Milk Coffee	Halib Qahua

Appendix B

The graphs shows individual participant performance on post-test immediately after training.