DIFFERENCE IN FOREIGN VOCABULARY LEARNING OUTCOMES BETWEEN VIRTUAL ENVIRONMENT IMMERSION-BASED, TEXT-BASED, AND PICTURE-BASED LEARNING

A Thesis

Presented to

The Faculty of the Department

of Psychology

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Arts

By

Brandin A. Munson

December, 2016

AN ABSTRACT OF A THESIS ON THE

DIFFERENCE IN FOREIGN VOCABULARY LEARNING OUTCOMES

BETWEEN VIRTUAL ENVIRONMENT IMMERSION-BASED, TEXT-BASED,

AND PICTURE-BASED LEARNING

A Thesis

Presented to

The Faculty of the Department

of Psychology

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Arts

By

Brandin A. Munson

December, 2016

ABSTRACT

The main focus of the present study was to compare foreign vocabulary learning outcomes between immersion-based, text-based, and picture-based training within a virtual environment. Researchers have yet to quantitatively compare outcomes of foreign vocabulary learning between students who use virtual environments as immersive tools and more traditional text-based and picture-based training methods. The present study explored differences across time between groups assigned to one of these training methods on quizzes testing generalization of foreign vocabulary to real-world pictures. A 3D virtual foreign vocabulary learning environment created by ESLI, a language learning company, was utilized. All three groups learned material within the game in order to minimize computer-based group differences, but only the immersion group was able to explore the world and see the physical objects, while the other two were limited to learning Spanish phonology from either 1) English text translations or 2) picture presentations, both within a classroom area. Each group completed all 3 sections of material on 4 separate occasions, and took quizzes on vocabulary knowledge and generalization after every section. Analyses were conducted on a final sample of 32 participants. T-tests revealed no differences between individual times, and immersion-based learning to have lower overall accuracy than either text-based or picture-based learning, with no significant differences between text-based or picture based learning. A 3x4 mixed-measures ANCOVA was conducted, comparing participants of different foreign vocabulary learning methods (either immersion, text, or picture) on quiz accuracy over a time period of 4 learning and testing sessions, while controlling for prior foreign (Spanish) vocabulary knowledge. While Spanish vocabulary and time each significantly predicted accuracy (p's < 0.05), neither the main effect of condition nor the interaction of condition and time significantly predicted accuracy (p's > 0.1). Implications and potential future directions are explored.

Terminology

- VE = Virtual Environment. Also defined in the text, this is a 3D videogame-like computer program which allows user movement and interaction with objects. This is distinct from virtual reality, which uses a headset. VEs only require control via a keyboard-and-mouse computer setup.
- Immersion = Unless otherwise specified, 'immersion' as used in the present text will refer to learning experiences which takes place within surroundings that are native to the object of study; in the case of the present study, in a location where learned words have sensory references which are directly experienced. This is distinct from classroom language immersion programs, where learners are exposed only to the target foreign language, often with large amounts of time dedicated to learning the target language.

Differences in Foreign Vocabulary Learning Outcomes Between Virtual Environment Immersion-based, Text-based, and Picture-based Learners

There are cognitive / health and academic benefits to learning a foreign language. For instance, some studies have associated bilingualism with a decrease in age-related losses in executive processing (Bialystok, Craik, Klein, & Viswanathan, 2004). Foreign language learning in high school has also been related to higher academic performance (Wiley, 1985) and higher scores on the verbal SAT and ACT (Cooper, 1987).

Foreign Vocabulary Learning

Much of the early time spent learning a foreign language is used on basic vocabulary acquisition. This often involves associating thousands of unique words with their conceptual referents, and can take learners anywhere from 500 to 2000+ hours to gain a moderate foreign language vocabulary (Schmitt, 2008). Because of the importance of early vocabulary acquisition in determining the degree of success in learning a foreign language, researchers remain interested in determining best practices for learning foreign vocabulary in optimal modalities, such as picture and text. However, findings remain mixed among investigations into the optimal methods for foreign vocabulary acquisition.

Lotto & de Groot (1998) compared foreign vocabulary learning outcomes among adult Dutch learners of Italian, between participants who learned word meanings either through 1) text presentation of the Italian word and the corresponding Dutch word, or 2) picture presentation alongside the corresponding Italian word. Participants were asked to *produce* (rather than identify, through reception) the corresponding word when presented with a word encountered, through text or picture, in the learning phase. Interestingly, for correct responses from participants and regardless of word cognate status, performance was better in the text condition than in the picture condition. However, other researchers conducting a similarly designed study have found alternative results.

For instance, Tonzar, Lotto & Job (2009) designed a study to determine modality differences in picture and text word learning, among Italian-speaking middle school learners of English and German. Similarly to Lotto & de Groot, participants learned word meanings either through 1) text presentation of the target word and the corresponding Italian word, or 2) picture presentation alongside the corresponding Italian word. Participants were tested on accuracy of identification after a delay following learning, and significant differences for learning condition, regardless of word cognate status, were found: those who learned through the use of picture-word pairings responded more accurately than those who used text-word pairings. This supports the authors' hypotheses, which drew from theories suggesting that a direct conceptual link is created when using pictures, while a secondary link to the L1 item is necessary when using text translations.

Modality of foreign vocabulary learning is especially important to consider in regions where student interaction with a language is particularly limited to the classroom setting. A large amount of foreign (that is, non-English) language learning in the United States occurs in classroom settings. According to the National Council of State Supervisors for Languages (NCSSFL), the majority of states in the U.S. require students to take some form of foreign language course for middle or high school graduation. However, these courses often rely primarily upon exercises involving first language translations into and from the foreign language. This overreliance upon a 'detached' method of learning, in which the learner is separated from the context in which the knowledge would normally be applied, has been criticized from a constructivist learning perspective (e.g., Gee, 2004).

Situated Cognition and Sensory Learning

Constructivist learning theories suggest that learners 'construct' learned models of the world based on the context in which novel material was experienced, called 'situated cognition' (Brown, Collins, Duguid, 1989; Fosnot & Perry, 1996). This fits with a number of sensory-based acquisition theories, including the sensorimotor learning hypothesis (e.g., Howell, Jankowicz, Becker, 2005; Hernandez & Li, 2007) and multisensory-emplaced learning (Fors, Backstrom & Pink, 2013). During language learning, according to sensory acquisition theories, the meanings and uses of words become connected to learners' embodied experiences while encountering those words. Thus, more diverse learning experiences result in more extensive neural activation during learning, with less reliance upon first language knowledge and leading to a more direct encoding of the learned language. Context-dependent memory, where learned material is better recalled in contexts similar to those in which it was learned, has been found in a wide range of studies (for review, see Smith & Vela, 2001). Such findings support the idea that language learning within a relevant environment context provides learners with knowledge more easily applicable to a real-world setting, where it would actually need to be used.

For these reasons, each of the above learning theories suggest that environmental immersion is more effective than learning that takes place separated from surroundings which are pertinent to the field of training, and studies comparing language learning environments have found certain advantages for more immersive learning. In a series of studies looking at language gains among groups of learners who differed in type of acquisition, researchers compared students who traveled to the destination where the target language was spoken with those who learned in a native classroom setting. The authors demonstrated that study abroad language learners show greater gains than classroom learners on behavioral measures such as oral fluency and amount of time spent speaking during conversation turns (Freed, Segalowitz, 2004; Segalowitz, Freed, 2004).

These findings are likely related to how memories are better encoded within more implicit learning situations. In a study comparing explicit artificial language learning with implicit learning, as a controlled proxy for a comparison of classroom with immersive learning, researchers did not find performance differences but did find that ERP activity related to proficiency and response to syntactic violations differed between groups (Morgan-Short, Steinhauer, Sanz, & Ullman, 2012). The authors concluded by saying that only implicit learning lead to native-like ERP activity, suggesting that "both the declarative and procedural memory brain systems can learn probabilistic patterns and to similar levels of performance, but only certain training conditions, in particular those in which explicit knowledge is minimized, lead to a processing dependence on procedural memory" (p. 944). It should be noted that in a follow-up study, the authors qualified this finding by showing that some learners respond differently to implicit stimuli than others (Morgan-Short et al., 2015). Future studies, including extensions of data gathered during the present investigation, ought to further investigate potential individual differences that could lead to observed differences in learning.

A direct comparison of immersion and classroom translation-based learning has been conducted, expanding experimental findings on explicit vs. implicit learning to a more applied context. Linck, Kroll & Sunderman (2009) compared classroom Spanish learners with those who studied abroad in Spain. Like Segalowitz & Freed (2004), the authors found that immersed learners outperformed classroom learners on comprehension and production tasks. However, the authors also found that immersed learners showed increased native language inhibition relative to classroom learners when presented with lexical-neighbor distractors while translating. This suggests that decreased native language use in the immersion context leads to decreased overlap between the learned and native languages.

It ought to be noted that comparisons of classroom and immersion learning often involves a number of factors other than level of immersion, including overall time spent learning and using the target language, and necessity of learning to fulfilling everyday needs within the foreign language context. Experimental testing of implicit vs. explicit learning is needed in order to more directly test whether these effects are due to levels of implicit learning. Nonetheless, taken together, a relevant learning environment and more implicit learning that does not rely on a native language seems to help students to become more proficient through the creation of sensory-based and not native language-based memories of foreign language material.

However, it remains difficult for American language learners to take advantage of immersion within a foreign language environment. Many U.S. states are geographically isolated from countries where languages other than English are predominantly spoken *and* English remains the world's lingua franca. Thus, not only is it geographically difficult for Americans to become immersed in a language other than English, but in business, academic, internet and media spheres there is decreased motivation for fluent English speakers to learn a foreign language (Lambert, 1989; Hornberger, 2002). A potential solution to these problems has been put forth, however, as teachers become increasingly involved in the use of more sophisticated technologies for language learning.

Advances in Vocabulary Learning Technology

As technology has advanced, so have its uses towards language learning, and vocabulary learning more specifically. Researchers have shown how novel word learning utilizing visual stimuli such as videos or pictures can be more effective than with text- or aural-based learning alone (e.g., Chung, 1994; Plass & Jones, 2005; Chun & Plass, 1996; and Al-Seghayer, 2001). In fact, a meta-analysis of the utility of different media towards language learning demonstrated large effects for vocabulary learning while being supported by a computer (Abraham, 2008).

In a study investigating the effects of different types of multimedia aides to learning, Yanguas (2009) compared Spanish vocabulary learning accompanied by text only, pictures only, or both text and pictures. Following a learning phase where participants learned Spanish vocabulary with one of these conditions, they were given immediate and delayed tests of vocabulary recognition. The tests of recognition involved textual presentation of a Spanish word, and selection of the correct English equivalent from a textual multiple-choice format. Participants whose Spanish learning was accompanied by both pictures and text outperformed each of the other groups in the delayed comprehension test, suggesting that multimodal presentation of learned words created meanings which lasted longer than unimodal presentations. Taken together, each of the above studies show how language learning can be improved when supported by visual and auditory stimuli.

These benefits are not limited to pictures and videos. A number of studies, detailed below, have shown that language learning is also supported where virtual environments are used. For the purposes of this study, a virtual environment (VE) will be defined as a 3D videogame-like computer program which allows user movement and interaction with objects. It is important to distinguish this from virtual reality, which requires the use of a headset in order to create a 360° first-person immersive experience. VEs, on the other hand, are used through traditional mouse-and-keyboard controls, much like a computer game.

Virtual Language Learning Environments

VEs provide a way to implement accessible situated cognition, giving learners an opportunity for an immersive language learning experience without the need of traveling to a region in which the language is spoken (Katelhut, Dede, Clarke, Nelson & Bowman, 2007; Dede, 2009; Mroz, 2012). Learner affordances enabled through VEs include a realistic visual environment, spatial and phonological feedback, embodied verbal and nonverbal communication, and embodied actions and object manipulations (Dalgarno & Lee, 2010). In a recent review of studies that have investigated language learning in a VE, Mroz (2014) concluded by saying that collaboration in creating VEs “should be the key for the L2 classroom to lose its walls and provide L2 learners with valid and efficient immersive experiences that have never so closely approximated real-life immersive learning in the target culture” (p. 341).

Studies have found that level of immersion within a simulation affects the amount of transfer that occurs during training- that is, the degree to which training in simulations generalizes to real-world applications (Alexander, Brunye, Sidman & Weil, 2005). This has also been shown to be a quantifiable variable- researchers point to both questionnaires and behavioral measures which adequately measure the level of learner’s immersion within a learning environment, called ‘presence’ (Jennett et al., 2008; Lee, 2004; Lombard & Ditton, 1997). Researchers have found levels of presence while playing video games (e.g., Yee, 2006) and during virtual learning (Baylor, 2009; Riva et al., 2007; Baker, 2010) to be related to player motivation. Among learners, motivation is strongly tied to learning outcomes both outside (Masgoret & Gardner, 2003) and inside (Yang, 2012) of VEs.

Currently, few studies investigating language learning in a VE have tested their use with each learner playing individually. Single-player learning standardizes the input each player receives by removing the social aspects of game interaction, and allows investigators to quantitatively asses the amount and type of language input players receive. To the present researcher's knowledge, no published study has conducted an experimental comparison of immersive VE language learning with more traditional learning methods inside or outside of a VE. Expanding upon suggestions made by Alexander, Brunye, Sidman & Weil (2005), one of the primary interests of the present study is testing vocabulary accuracy and transfer to real-world situations: as they learn, do VE immersion vs. Text vs. Picture learners differ on how well they are able to retain material and transfer to real-world situations?

The aim of the present study is to address these issues by comparing language learners who hear Spanish nouns auditorily while interacting with 3D objects in a simulation of a small Spanish town with learners who remain in a simulated school setting associating auditorily-presented Spanish nouns with either 1) English text translations or 2) picture presentations. All 3 groups learned in an online VE setting in order to account for potential confounds associated with the degree of complexity in using computer technology, but only the environmental immersion group explored the town setting and interacted with 3D representations of learned objects. Meanwhile, text-based and picture-based learning groups only encountered Spanish material via Spanish phonology and either English text or nonverbal pictures. All groups took quizzes modeled similarly to those used in Yanguas (2009), which tested knowledge of Spanish with real-life pictures in order to measure the transfer of content learned in each learning condition. Accuracy was measured across 4 sessions of play in order to assess longitudinal differences in language learning for each of the groups.

Hypotheses

1) Regardless of condition, accuracy on quizzes of Spanish vocabulary was expected to increase over each of the 4 time points. Using pairwise t-tests, later time points were expected to have higher accuracy than earlier time points.

2) In accordance with qualitative trends from past VE language learning studies (see Mroz, 2014 for review), higher accuracy was expected in the immersion-based condition relative to each of the other conditions, without taking time into account. Using pairwise t-tests, immersion learning was expected to have higher accuracy than 1) text learning and 2) picture learning, and text learning and picture learning were expected to not differ.

3) Immersion-based learners were expected to show earlier increases in accuracy over the 4 sessions than either text-based or picture-based learners. More specifically, with overall quiz accuracy as the dependent measure for each of the following hypotheses and while controlling for basic Spanish vocabulary knowledge, an interaction effect between learning environment condition (immersion, text and picture) and time (4 sessions) was expected to be found through a 3x4 mixed measures ANCOVA (see “Data Analysis” section).

Method

Participants

Eighty participants completed the first in-lab portion of the study, and of the 80, 12 (15 percent) finished at least 1 but not all 4 sessions, and 28 (35 percent) finished all 4 sessions of virtual environment. Of those who finished at least 1 session, 6 were found to have 3 or more years of high school Spanish classroom experience, and 2 experienced in-game glitches that resulted in going through off-limit areas of the environment early on. Data from these 8 participants was removed. All other participants were monolingual English speakers, had 2 years or less of formal high school Spanish learning, and no formal college Spanish learning experience. Thus, data from 32 participants (22 female) was used for the final analyses.

All participants were randomly assigned to each of the three conditions: Immersion learning, Text learning, and Picture learning. Of the 32 participants used in analyses, 14 were in the immersion condition, 10 were in the text condition, and 8 were in the picture condition. See Table 1 for descriptive demographic information. All participants were English-speaking monolinguals recruited from the UH main campus. Those who completed all parts of the study received 6 SONA credits as compensation.

Though the overall number of participants brought to the lab (80) was higher than the expected number of 60 necessary to reach a minimal level of power, retention for completion of the first learning session (50%) was below that of the pilot study (70%). See the ‘limitations’ section for the influence of this on interpretation of the present study’s analyses, with regards to power assumptions.

Screening

Collection of preliminary screening and questionnaire forms took place at the University of Houston, in the Laboratory for the Neural Bases of Bilingualism (LNBB). Here, participants filled out a lab-constructed background questionnaire, detailing how much (if any) foreign language experience they have, as well as musical experience, gaming experience, computer attitudes, and certain socioeconomic and demographic information. Those who reported either a) being proficient in 2 or more languages, or b) having taken more than 2 years of high school, or any college Spanish classes, were not included in the present analyses. Participants also completed the picture and vocabulary comprehension subtests of the Woodcock Language Proficiency Battery-revised, English Form (Woodcock & Munoz-Sandoval, 1995), as well as a basic vocabulary test of Spanish proficiency, with no overlap with in-game vocabulary. All of this information was included during multiple imputation of the current data (see 'analyses' below for more information), and will be used in future investigations on the effects of individual differences and in-game behavior towards outcomes. All of the information collected has been shown to influence learning effects and / or interaction with computers past studies; e.g., music experience has been found to be related to phonological perceptual abilities (Bolduc, 2009), and gaming experience determines and computer attitudes each influence efficacy and motivation in using video games to learn (Orvis, Orvis, Belanich & Mullen, 2005).

Procedure

Study Structure. The study took place in 3 parts, with detailed descriptions of each aspect in sections below. Part one, which took place in the lab, consisted of initial screening, questionnaire completion, and introduction to the VE. Part 2, which was conducted online, consisted of participants logging into and playing the VE over 4 separate sessions from a personal computer outside of the lab. Part 3, which again took place in the lab, consisted of completion of a presence questionnaire and an exit survey.

Immersion Condition. Playing the VE in the immersion-based learning condition involved exploring a Spanish town square and home, finding and clicking sound icons near objects to hear their Spanish words, and taking a picture-word association quiz at the end of each section. For instance, learning the word 'dog' in the immersive learning condition involved clicking on the sound icon next to a 3D representation of a dog, and hearing the Spanish word 'el perro' in response. There were 3 sections overall with 40 unique Spanish word / object pairs in each section. Following similarly designed studies comparing learning conditions, all words were concrete, picturable nouns (e.g. Lotto & de Groot, 1998), in order to minimize ambiguity when inferring meaning. Every section was followed by a unique picture-word association quiz (detailed below). Participants completed 4 sessions of play, each of which consisted of moving through each of the 3 game sections and taking all 3 quizzes. Each session was completed at least 1 day and at most 2 days apart, meaning that all 4 sessions were completed in at least 4 days, and at most 8 days.

As participants played, data such as object interactions, time spent within each section and quiz answers were automatically sent to an encrypted server for researcher use.

Text and Picture Conditions. The text-based and picture-based learning conditions involved using the same Singularity computer program, but participants were restricted to a single classroom within a school setting and seeing English text translations (text condition) or pictures (picture condition) when hearing Spanish words via audio, rather than exploring each of the 3 sections and seeing the 3D representations of objects. Participants in the text and picture conditions went through material for each of the same 40 words in each section in the same order as those in the immersive learning condition, with the option to repeat material when desired. The same quizzes were used as in the immersion learning condition.

The reason the text and picture conditions also took place within the VE is twofold. First, this created consistency in the learning experience across participants, so that all groups differed only with respect to the methodology used during language training- one group explored and interacted with 3D objects while hearing Spanish words, and the others viewed and moved forward or backward through PowerPoint-like presentation of text or pictures while hearing Spanish words. The second reason was that this allowed for the same in-game interaction, time and accuracy data to be automatically collected from all groups of participants.

Stimuli

Virtual Environment Program. The VE was created by the company 'English as a Second Language International' (ESLI) with input from the present researcher. ESLI specializes in the instruction of foreign languages utilizing novel language acquisition materials. The VE runs off of a program called 'Singularity,' which is the same engine that runs the popular and widely researched online game 'Second Life.' Playing of the Spanish words, quiz administration, automatic data collection and the player information display were enabled by coding a heads-up display (HUD) within Singularity. With this HUD, players can interact with objects to hear Spanish words, see the number of words in a section, keep track of how many are left to find before taking the quiz, and are given the quiz once all words have been found.

Spanish Word Recordings. In-game Spanish phonology was recorded by a native Spanish speaker in a noise-eliminating sound booth in the University of Houston Center for Communication Sciences and Disorders. Recordings included 120 non-cognate Spanish nouns selected from the 'Puntos' and 'Connectados' Spanish textbooks. Praat (Boersma, 2001) was used to cut each sound file and to normalize and enhance the recordings. Spanish word recordings were uploaded to Singularity, and the HUD was updated so that sound files would play as participants interacted with certain objects.

VE Quizzes. Quiz questions consisted of 6 visual items presented simultaneously during the auditory playing of a Spanish word previously learned in the environment or classroom sections. Only one visual item out of the six matched the heard Spanish word. Items were either 1) real-world pictures which depicted objects learned in-game or 2) English textual translations of Spanish words for objects learned in-game. Being quizzed on the word 'dog,' for instance, would involve one of the 6 items appearing as either the real-world picture presentation of a dog or the English text presentation 'dog.' Participants might hear the Spanish word 'perro,' and the correct answer for that question would be to click on the picture or text item associated with 'dog.' All 6 quiz items were either presented simultaneously as either pictures or text, and only one representation of each word could occur for any individual quiz item.

The quiz structure across each of the 4 sessions is displayed in Figures 1 (novel vs. repeated items) and 2 (picture vs. text items) in the Appendix. For the first session of play, each of the 3 section quizzes consisted of 10 novel object questions. For the 2nd through 4th sessions of play, each of the 3 section quizzes consisted of 10 novel questions and 10 questions repeated from the quiz taken during the session prior. This allowed for novel quiz questions throughout all 4 sessions, giving us a measure of pure environment foreign vocabulary learning regardless of past quizzes through all 4 sessions, and for the control and potential testing effect of quiz material repetition in sessions 2 through 4. This also ensured that participants did not begin to ignore distractor pictures of objects which they may have already seen. In all, 4 in 5 quiz questions are presented as object pictures, and 1 in 5 as English text. Text questions will clarify, in future analyses, whether use of immersion learning differs from textual learning when translating back to the native language. A random number generator was used to select and order items for each quiz, which items were presented as pictures or text, and which items were repeated across quizzes. All quiz items used in present analyses were novel questions, so that effects of in-game learning prior to each quiz were the primary driver of responses, and not feedback from earlier quiz answers. By using novel questions only, in-game exposure to words is not affected by quizzes, but only by how much participants interacted with objects of their own volition. Repeat quiz questions will be used in future analyses to test the effects of quizzes on learning.

This quiz structure, implemented with input from the VE designer and coder, allowed for a direct comparison of the generalizability of immersion-based vs. text-based vs. picture-based learning to real-world objects.

Picture Selection and Presentation. Two-hundred and forty free-use pictures (80 per section) were collected for quizzes by using 1) advanced Google image searching and 2) the Pixabay non-commercial image database. Using Adobe Photoshop, pictures were resized to 256x256 pixels and then uploaded into Singularity, and the HUD was updated so that pictures of correct answers matched the random order of each quiz. This order was predetermined from randomly generated number lists for 1) ordering of items through the 4 sessions of each environment section, 2) determining which items were presented as pictures or text, and 3) determining which items were repeated from one quiz into the next. Pictures displaying objects that are different from the correct picture were selected randomly from the complete list of 80 section pictures. Each object was represented by two pictures because repetitions of correct answers cannot use the same picture, and also because it increases the diversity of the distractor pictures.

One-hundred and twenty additional free-use pictures (40 per section) were also collected for the picture-based learning condition by using 1) advanced Google image searching and 2) the Pixabay non-commercial image database. Using Adobe Photoshop, pictures were resized to 256x256 pixels and then uploaded into Singularity, using the same design as those in text-based learning, with the single exception that English words were replaced with single pictures.

Data Analysis

Forty participants completed part 1 of the study, as well as at least 1 in-game session, qualifying their data for analyses. Participants who reported more than 2 years of formal high school, or any college, Spanish experience were discarded (for a total of 6 discarded participants). Participants who experienced glitches where they played sections of the game normally unavailable were also discarded (for a total of 2 participants). Therefore, data from a total 32 participants was used in these analyses.

Multiple Imputation. Of those 32 participants, 23 completed all 4 in-game sessions, while 9 completed at least 1 full in-game session, but not all 4. Data was therefore multiply imputed using the MICE package (van Buuren & Groothuis-Oudshoorn, 2011) in the R statistical software (R Core Team, 2016) in order to make use of every possible participants' data. Multiple imputation, a process whereby missing data is simulated over several imputations through the use of all relevant non-missing variables for all participants, has been shown to be able to estimate values for missing data points while maintaining standard errors and overall variability within data (e.g., Rubin, 1996).

During multiple imputation, any variables which may influence variables of interest, whether outcomes or predictors, need to be included in the imputation structure in order to account for variance associated with those variables. These do not need to be included in subsequent analyses, but must include all variables for potential future analyses. Variables included in the multiple imputation process were: 1) age, 2) factored condition variable (immersion, text or picture), 3) Spanish vocabulary knowledge, 4) English proficiency as measured by the Woodcock language proficiency battery-revised (Woodcock & Munoz-Sandoval, 1995), 5) SES as an average of parent's occupations, 6) self-rating of motivation to learn Spanish, 7) number of reported familiar languages, 8) college GPA, 9) college year (e.g. Freshman, Sophomore), 10) gender, 11) mother's education, 12) father's education 13) motivation to learn any foreign language, 14) past usage of Rosetta Stone, or other language-learning software, 15) hours spent playing video games on a workday, 16) hours spent playing video games on a weekend, 17) familiarity with 3D video games, 18) familiarity with 2D video games, 19) extent of prior video game usage, 20) extent of current video game usage, 21) ownership of a computer capable of playing 3D video games, 22) enjoyment of the idea of learning through a video game, 23) enjoyment of explicit learning, rather than implicit learning, 24) accuracy score for time 1 for each participant, 25) accuracy score for time 2 for each participant, 26) accuracy score for time 3 for each participant, 27) accuracy score for time 4 for each participant.

Fifty iterations were created, and the average of 5 randomly selected iterations was taken in order to reshape the data into a format necessary for conducting a mixed-measures analysis of covariance (ANCOVA). Analyses were conducted on each of the five iterations, and averages of these were taken for the final results.

Results

Hypothesis 1. The first hypothesis, which stated that accuracy would increase over time regardless of condition, was not supported. Although accuracy did slightly increase through Time 1 ($M = 70.31\%$, $SD = 16.52$), Time 2 ($M = 72.53\%$, $SD = 13.5$), and Time 3 ($M = 79.8\%$, $SD = 13.51$), and did not change through Time 4 ($M = 79.75\%$, $SD = 11.69$), none of the comparisons between times were significant after using the Bonferroni correction for multiple comparisons (see Table 2). However, two comparisons did closely approach significance: first, Time 1 increasing through Time 3, $t(59.66) = -2.51$, $p = 0.088$; and second, Time 1 increasing through Time 4, $t(55.8) = -2.64$, $p = 0.066$. This lack of clear significance is likely due to a low overall N; see Limitations section below.

Hypothesis 2. The hypothesis that the immersion condition would show higher overall accuracy than the other conditions regardless of time was not supported, and in fact, the opposite was observed. Pairwise t-tests were conducted, comparing each condition's (immersion, text and picture) overall accuracy regardless of time (see Table 3). The Bonferroni correction for multiple comparisons was used. Participants in the immersion-based learning condition (M = 71.54%, SD = 14.62) were not significantly lower in overall accuracy than participants in the text-based learning condition (M = 78.35%, SD = 13.63), though the comparison very nearly approached significance at $t(87.54) = -2.34$, $p = 0.06$. Participants in the immersion-based learning condition were significantly lower in overall accuracy than participants in the picture-based learning condition (mean accuracy = 79.27%, SD = 13.51), $t(69) = -2.51$, $p = 0.044$. Participants in the text-based learning condition did not differ from participants in the picture-based learning condition, $t(66.82) = -0.28$, $p > 0.05$.

Hypothesis 3. Hypothesis 3 stated that as time passed during learning and quiz-taking, accuracy would increase at different rates depending on participant condition. See Table 4 for mean values and SD's of each condition at each time point, and Figure 3 for a bar plot graphing differences in accuracy for each color-coded condition across time. A single ANCOVA was estimated in the R statistical software on all 32 participants. Hypothesis 3 was not supported by the model output. The model outcome variable was quiz accuracy, and predictors were 1) Spanish vocabulary knowledge, 2) time, 3) condition, and 4) interaction of time and condition. The covariate of prior Spanish vocabulary knowledge was significant, $F(2, 1) = 12.46$, $p = 0.007$. The high significance of prior Spanish vocabulary knowledge demonstrates the importance of its inclusion as a covariate. Quiz accuracy did not change at different rates over time depending on condition when controlling for Spanish vocabulary knowledge, $F(2, 1) = 1.658$, $p = 0.15$.

Because the interaction term was found to be nonsignificant (see the Limitations section for discussion about the low overall N), the main effects of the predictor variables become informative. It was assumed that 1) as time passed, accuracy would differ, and 2) that accuracy would differ between conditions. The main effect of time was significant, $F(3, 6) = 6.77$, $p < 0.001$. So, accounting for variability due to condition, Spanish vocabulary and the interaction between condition and time, accuracy was still very clearly seen to differ over time. For directions of the change over time, see Hypothesis 2 (above). The main effect of condition was nonsignificant, $F(3, 6) = 2.3$, $p = 0.12$. So, accounting for variability due to the other predictors, accuracy did not significantly differ between the three conditions.

Discussion

The purpose of this experiment was to determine differences in foreign vocabulary outcomes among those who learn through 3D immersion-based methods, those who learn through text-based methods, and those who learn through picture-based methods, all within a virtual environment. The main hypothesis of interest, that immersion learning would cause faster gains over other learning methods, was not supported. Across times, immersion-based learning participants showed no differences, trending towards lower overall gains regardless of time, in accuracy for a task involving transferring learned vocabulary to novel pictures than did either text-based or picture-based learners. The importance of these findings for the use of technology in language learning is considered, and the limitations of the current experiment and potential future directions are explored.

The results from the present study suggest that, on a task of transferring learned vocabulary words, concrete nouns, to novel pictures representing the meanings of those words, immersion-based learning does not result in differences between either text-based or picture-based learning, and may in fact trend in the direction of being less beneficial towards learning. Because this is the only study to date, to the researcher's knowledge, directly comparing the benefits of 3D vocabulary learning with more traditional vocabulary learning, replications and extensions are needed in order to confirm and qualify the present findings. Nonetheless, taking the present findings into consideration, it is possible that claims about the efficacy of VEs towards language learning have been overstated in the past literature.

The lack of benefits found among learners in the immersion condition relative to the other learning conditions was unexpected. Although this experiment consisted of a novel comparative approach to testing differences in outcomes among immersion and more traditional (text, picture) 3D language learning approaches, a sizable qualitative literature suggests motivational and sensorimotor benefits for those who learn within 3D environments (see Mroz, 2014 for review). These findings to the contrary, demonstrating less accuracy in the immersion condition, could potentially be explained through a differing test framework.

There is the possibility is that the time separating learning and testing may have been too short to determine differences in long-term memory retention between learning methods. Although the ANCOVA found no accuracy differences between conditions, this finding is limited to the timeframe in which testing was conducted- closely following learning sessions, rather than after a significant delay. Other studies which have explored memory differences between learning methods have used a larger time delay between learning and post-tests, such as Yanguas (2009), who tested differences in vocabulary retention between those who used text and those who used pictures. Yanguas tested participant's knowledge roughly 3 weeks following learning, and found that pictures aided word learning relative to text alone, a finding that was not replicated in the present study. Such a delay could potentially have drawn out larger differences in memory decay between groups, rather than testing shorter-term (though still not working) memory and the ability to generalize to novel pictures. Future studies investigating memory differences should use a larger amount of time following learning before memory is tested.

Limitations and Future Directions

An important limitation in the present study relates to the low N collected for the final sample. Although 80 participants were able to come into the lab for Part 1 of the study, less than half (32) fully completed the first in-game section, and only 23 completed all in-game sections. This was lower than the retention rate during the pilot study, which was roughly 50%. This could be due to 1) use of only SONA credit, and not gift cards, following the pilot study, or 2) a decrease in the pool of motivated potential participants who were monolingual, without much Spanish language experience, and signed up in SONA following pilot data collection. Because of this, and drawing from effect sizes observed in similar past studies, the observed interaction effect likely does not have a large enough N to be conclusive. In a study comparing combinations of vocabulary learning glosses (phonology, text, and pictures) Kim & Gilman (2008) found a large ANOVA main effect for differences across learning methods (5 groups; N = 172; effect size $\eta^2 = .826$), but a small interaction effect between learning conditions across time ($\eta^2 = .079$). This suggests that more participants will need to be collected before sufficient power for determining an interaction effect from the present study is possible. A larger N would also allow for more flexible analyses and potentially influential covariates to be added, such as amount of interaction during learning and time delay between learning sessions. Data collection is ongoing for this purpose.

One potentially confounding variable is the timing of picture-based learning relative to other learning conditions. Whereas immersion-based and text-based learning conditions prepared and collection was ongoing early in Spring 2016, picture-based learning was not prepared until late Spring 2016. Although data was randomly collected otherwise, this could have led to an implicit bias in collection. For instance, although retention among picture-based learners was higher, this was likely because a larger proportion of data collected among those in this condition occurred later in the Spring semester than did for the other conditions.

Also, the VE remains in an early state of completion, and currently only involves foreign vocabulary learning with 120 in-game words, and the present experiment was thus limited to this portion of the target language. Expansion of the physical environment and Spanish vocabulary included, as well as the grammar learned in-game, is currently under construction. Such expansions will allow for more extensive, and possibly complex, vocabulary learning to be tested. Expansion to other aspects of language, such as grammar, could also allow for a more generalizable testing of VEs towards language learning, rather than foreign vocabulary learning specifically.

Contact with high school language teachers is also underway, and the hope is that this will lead to a more applicable test of the learned material. High school students are closer to the target age for VEs, and are less likely to have significant Spanish language experience. Also, learning in the VE can then be compared with other types of supplemental classroom learning, over a longer length of time, leading to more robust measures of potential differences.

Finally, participants will be contacted for additional testing to investigate long-term retention of learned vocabulary. One possibility explored in the Discussion section above is that the timing of tests did not allow for potential memory decay effects to arise. These additional tests will help to clarify whether longer durations between learning and testing lead to different learning effects. Future participants will also be given a later post-test in order to determine differences in longer term memory of vocabulary between conditions.

There are many improvements that can be made upon this study, but a step forward has nevertheless been taken in creating an experimental comparison of 3D immersion-based learning of foreign vocabulary relative to more traditional text-based and picture-based for the first time. Although the hypotheses suggesting benefits of immersive virtual learning were not supported, this is nevertheless informative to teachers by limiting some of the potentially overstated excitement behind making use of certain technologies to aid student learning. Although the potential for learning benefits in new technologies remain, the present study clearly demonstrates the need for more formal investigations into how and why novel forms of learning, including VEs, might actually be more beneficial than more traditional learning methods before large claims are made. Only then will these methods be able to be incorporated into an actual classroom or online learning context in an effective manner.

References

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, *21*(3), 199-226.

Alexander, A. L., Brunyé, T., Sidman, J., & Weil, S. A. (2005). From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group*, *5*, 1-14.

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, *5*(1), 202-232.

Baker, C. (2010). The Impact of Instructor Immediacy and Presence for Online Student Affective Learning, Cognition, and Motivation. *Journal of Educators Online*, *7*(1), 1-30.

Baylor, A. L. (2009). Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1535), 3559-3565.

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and aging*, *19*(2), 290.

Boersma, P., & Weenink, D. (2001). Praat 3.9. 15 [Computer software].*Amsterdam, the Netherlands: Institute of Phonetic Sciences*.

Bolduc, J. (2009). Effects of a music programme on kindergartners' phonological awareness skills 1. *International Journal of Music Education*, *27*(1), 37-47.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational researcher*, *18*(1), 32-42.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The modern language journal*, *80*(2), 183-198.

Chung, U. K. (1994). *The effect of audio, a single picture, multiple pictures, or video on second-language listening comprehension* (Doctoral dissertation, University of Illinois at Urbana-Champaign).

Dalgarno, B., & Lee, M. J. (2010). What are the learning affordances of 3-D virtual environments?. *British Journal of Educational Technology*, *41*(1), 10-32.

Dede, C. (2005). Planning for “neomillennial” learning styles: Implications for investments in technology and faculty.

Dede, C. (2009). Immersive interfaces for engagement and learning. *science*,*323*(5910), 66-69.

Cooper, T. C. (1987). Foreign language study and SAT-verbal scores. *The Modern Language Journal*, *71*(4), 381-387.

Fors, V., Bäckström, Å., & Pink, S. (2013). Multisensory emplaced learning: Resituating situated learning in a moving world. *Mind, Culture, and Activity*,*20*(2), 170-183.

Fosnot, C. T., & Perry, R. S. (1996). Constructivism: A psychological theory of learning. *Constructivism: Theory, perspectives, and practice*, 8-33.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*(02), 275-301.

Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. Psychology Press.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: its neural and computational mechanisms. *Psychological bulletin*, *133*(4), 638.

Hornberger, N. H. (2002). Multilingual language policies and the continua of biliteracy: An ecological approach. *Language policy*, *1*(1), 27-51.

Howell, S. R., Jankowicz, D., & Becker, S. (2005). A model of grounded language acquisition: Sensorimotor features improve lexical and grammatical learning. *Journal of Memory and Language*, *53*(2), 258-276.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, *66*(9), 641-661.

Ketelhut, D. J., Dede, C., Clarke, J., Nelson, B., & Bowman, C. (2007). Studying situated learning in a multi-user virtual environment. *Assessment of problem solving using simulations*, 37-58.

Kim, D., & Gilman, D. A. (2008). Effects of Text, Audio, and Graphic Aids in Multimedia Instruction for Vocabulary Learning. *Educational Technology & Society*, *11*(3), 114-126.

Lambert, R. (1989). The national foreign language system. *CALICO Journal*,*6*(4), 7-22.

Lee, K. M. (2004). Presence, explicated. *Communication theory*, *14*(1), 27-50.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*(12), 1507-1515.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer‐Mediated Communication*, *3*(2).

Lotto, L., & De Groot, A. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language learning*, *48*(1), 31-69.

Masgoret, A. M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: a meta–analysis of studies conducted by Gardner and associates. *Language learning*, *53*(1), 123-163.

Morgan-Short, K., Deng, Z., Brill-Schuetz, K. A., Faretta-Stutenberg, M., Wong, P. C., & Wong, F. C. (2015). A view of the neural representation of second language syntax through artificial language learning under implicit contexts of exposure. *Studies in Second Language Acquisition*, *37*(02), 383-419.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of cognitive neuroscience*, *24*(4), 933-947.

Mroz, A. P. (2012). Nature of L2 negotiation and co-construction of meaning in a problem-based virtual learning environment: a mixed methods study.

Mroz, A. (2014). 21st Century Virtual Language Learning Environments (VLLEs). *Language and Linguistics Compass*, *8*(8), 330-343.

Orvis, K. A., Orvis, K. L., Belanich, J., & Mullin, L. N. (2005). *The influence of trainee gaming experience and computer self-efficacy on learner outcomes of videogame-based learning environments* (No. TR-1164). Army Research Inst for the Behavioral and Social Sciences Alexandria VA Research and Advanced Concepts Office.

Plass, J. L., & Jones, L. (2005). Multimedia learning in second language acquisition. *The Cambridge handbook of multimedia learning*, 467-488.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., & Alcañiz, M. (2007). Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior*, *10*(1), 45-56.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association*, *91*(434), 473-489.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language teaching research*, *12*(3), 329-363.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in second language acquisition*, *26*(02), 173-199.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review*, *8*(2), 203-220.

Tonzar, C., Lotto, L., & Job, R. (2009). L2 vocabulary acquisition in children: Effects of learning method and cognate status. *Language Learning*, *59*(3), 623-646.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. URL http://www.jstatsoft.org/v45/i03/.

Wiley, P. D. (1985). High School Foreign Language Study and College Academic Performance. *Classical Outlook*, *62*(2), 33-36.

Wong, C. W., Olafsson, V., Plank, M., Snider, J., Halgren, E., Poizner, H., & Liu, T. T. (2014). Resting-state fMRI activity predicts unsupervised learning and memory in an immersive virtual reality environment. *PloS one*, *9*(10).

Woodcock, R. W., & Muñoz-Sandoval, A. F. (1995). Woodcock language proficiency battery revised: English and Spanish forms. Riverside Publishing Company.

Yang, Y. T. C. (2012). Building virtual cities, inspiring intelligent citizens: Digital games for developing students' problem solving and learning motivation. *Computers & Education*, *59*(2), 365-377.

Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning & Technology*,*13*(2), 48-67.

Yee, N. (2006). Motivations for play in online games. *CyberPsychology & behavior*, *9*(6), 772-775.

Appendix

Table 1

Group means and standard deviations for participants averaged within each learning condition. Standard deviations are in parentheses.

	All Participants	Immersion	Text	Picture
N	32	14	10	8
Proportion Female	0.68	0.64	0.67	0.75
SES	14.43 (6.67)	13.32 (5.59)	14.73 (0.6)	16 (5.23)
GPA	3.1 (0.42)	3.11 (0.45)	3.13 (0.6)	3.05 (0.41)
Spanish Vocabulary Knowledge	0.54 (0.21)	0.5 (0.14)	0.65 (0.29)	0.46 (0.2)
English Proficiency	35.3 (8.16)	32.38 (6.18)	36.2 (4.8)	38 (5.93)
Motivation to Learn Spanish	3.5 (1.15)	3.64 (1.06)	3.3 (0.93)	3.5 (0.95)

Table 2

Means, standard deviations, and t values of observed accuracy averages for each time point, regardless of learning method. All comparisons nonsignificant after Bonferroni corrections.

	Mean	SD	Time 1	Time 2	Time 3	Time 4
Time 1	70.31	16.52	1	-0.58	-2.51	-2.64
Time 2	72.53	13.5		1	-2.15	-2.29
Time 3	79.8	13.51			1	0.015
Time 4	79.75	11.69				1

Table 3

Means, standard deviations, and t values of observed accuracy for each learning condition, regardless of time.

	Mean	SD	Immersion	Text	Picture
Immersion	71.54	14.62	1	-2.34	-2.51*
Text	78.53	13.63		1	-0.28
Picture	79.27	13.51			1

*=p<.05

Table 4

Means and Standard Deviations of accuracy for each condition across the 4 times. Standard Deviations are in parentheses.

	Immersion	Text	Picture
Time 1	71.25 (17.86)	70.25 (17.5)	68.75 (14.76)
Time 2	66.61 (14.29)	76.6 (10.89)	77.81 (12.21)
Time 3	74.21 (14.18)	84.95 (11.76)	83.12 (12.01)
Time 4	74.07 (11.79)	81.6 (10.24)	87.38 (8.71)

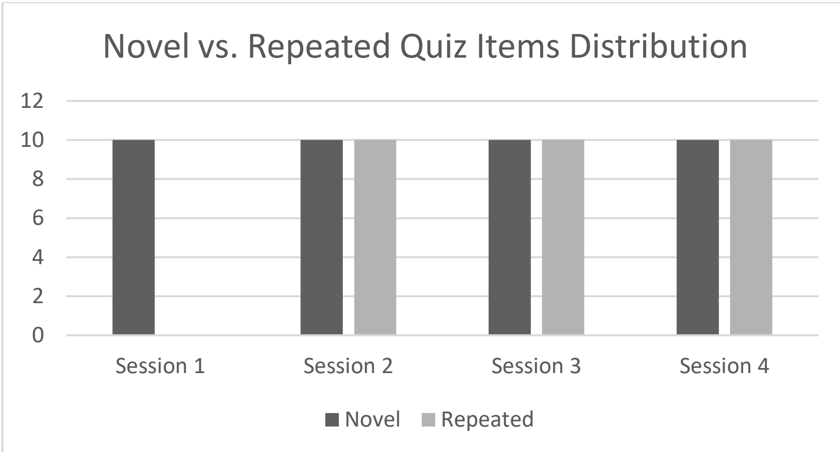

Figure 1. *Quiz question types across each of the 4 sessions. There are 10 quiz questions in the first session, and 20 in sessions 2, 3 and 4. Every session contains 10 novel quiz items, and sessions 2, 3 and 4 also contain 10 questions repeated randomly from the session prior. Picture and text questions are split evenly among novel and repeated questions.*

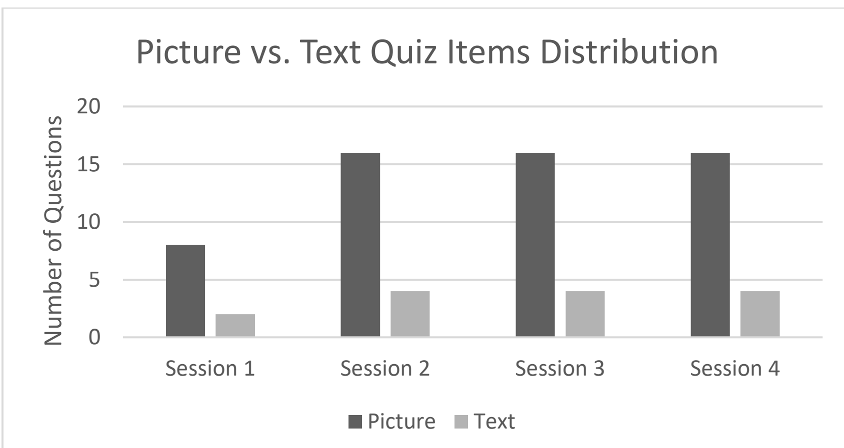

Figure 2. *Quiz question types across each of the 4 sessions. There are 10 quiz questions in the first session, and 20 in sessions 2, 3 and 4. One in five quiz questions are represented as text while the others are pictures. Picture and text questions are split evenly among novel and repeated questions.*

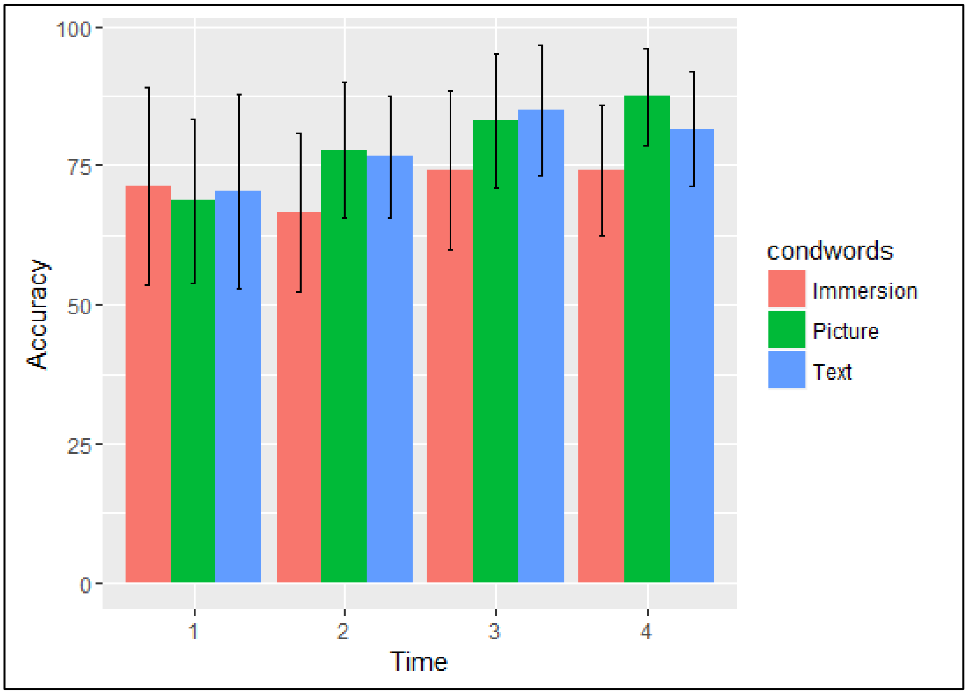

Figure 3. *Bar graph of mean accuracy for each condition at every time point. See Table 4 for mean values and SDs. Error bars depict one SD above and one SD below the mean at each time point for each condition.*