Second Language Acquisition and Memory Training Transfer: Could working memory training facilitate the learning of a second language?

Jessica Godes
University of Vermont

Second Language Acquisition and Memory Training Transfer:
Could working memory training facilitate the learning of a second language?

by Jessica Godes

Advised by Guillermo Rodríguez

University of Vermont
College of Arts and Sciences
Department of Romance Languages and Linguistics

April 17th, 2019

Acknowledgements

Thank you to the College of Arts and Sciences and the Department of Romance Languages and Linguistics for the funding obtained from the APLE Award.

Thank you to Professors Uzzell and DeDios for allowing us to come into their classrooms and explain our research to their Spanish 001 students, and for all of the work they do in the Spanish program at UVM.

Thank you to the student research assistants who helped me to gather research for this project: Annie Wernhoff, Katie White, Eve Pomazi, Emma Russo, and Jess Van Houten.

Thank you to Professors Swanson and VonDoepp for sharing their time and efforts on my thesis committee.

Thank you especially to Professor Guillermo Rodríguez. Without his support and encouragement, this project never would have gotten off the ground. Thank you for your endless source recommendations, email responses, editing sessions, and explanations of statistics.

Abstract

Working memory (WM) training has been reported to be effective in not only improving WM capacity, but in transferring to other cognitive domains. However, although recent studies are consistent in reporting improvement in the specific task used for training, not all skills seem to transfer. This study seeks to examine the potential for WM training transfer to other cognitive skills, particularly those used in the acquisition of a second language as an adult (such as fluid intelligence and the focus of attention). Participants were college students studying Spanish at a beginner level. They were split into three conditions: a control group which received no WM training; a target condition which underwent adaptive WM training; and an active control group who did a non-adaptive version of the task assigned to the target group. Due to the high percentage of attrition during the study, the results were examined longitudinally based on the training regimen of each participant. One participant who completed a high number of adaptive training sessions showed improvement in all four cognitive measures used in the pre- and post-test phases, as well as in the grammar test used to measure acquisition of Spanish grammar; however, their performance regarding grammar was not to the level expected if their WM training were to affect language acquisition. Other participants who completed some non-adaptive training sessions also showed improvement in cognitive measures, although the scale of their improvement does not appear to relate with the number of sessions they completed. The data show a positive correlation between WM training and improvement in various cognitive skills, but the relationship between cognitive training and foreign language grammar improvement is not clear cut.

1. Introduction

In this project, we attempted to determine if training working memory (WM) capacity will help young adults to learn a second language (L2) more efficiently. Previous research has shown that training one's WM will make it more efficient, showing evidence of transfer to tasks that rely on other cognitive capacities such as focusing attention (Lilienthal, Tamez, Shelton, Myerson, & Hale, 2013) or general fluid intelligence (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). We hypothesized that training WM will make it more efficient, and that the efficiency developed during training transfers to skills involving WM that are necessary for learning the grammar of a new language as an adult in a classroom setting.

The intention was to contribute to the body of knowledge about the phenomenon of transfer between WM training and other cognitive skills, such as the learning of a foreign language in an instructed setting. In gaining a greater understanding of the role WM plays in second language acquisition (SLA), this study could form a basis to rework strategies for teaching foreign languages based on WM capacity as an individual cognitive difference. If there were evidence of transfer between WM training and grammar knowledge, then WM training could be used as an additional tool to help students learn languages more efficiently.

Literature Review

1.1 Working Memory

WM serves as an intermediary between short-term and long-term memories. New information is processed in WM, held there for a short timespan before it can be transferred to long term memory (LTM) (Baddeley, 2012). Information is also recalled from LTM to be manipulated in WM, integrated with new or other old information. According to the most prevalent paradigm in the field, Baddeley's Multicomponent Model (Baddeley, 2012), WM is comprised of different components that achieve different functions. The Central Executive (CE) is responsible for delegating attention, and is made up of various executive functions (EFs). EFs are specific processes that come about as a result of focusing attention. Most relevant to us are updating, in which information is constantly being renewed and kept fresh within WM, and conflict resolution, where the cognitive system is able to solve temporary difficulties or ambiguities in language processing. WM capacity includes systems controlled by the CE that are used for visual processing (the visuospatial sketchpad) and auditory processing (the phonological loop). There is also an episodic buffer, where information from the various sensory inputs is combined and stored (i.e. putting together visual, auditory, and tactile memories to create one full image). This buffer not only connects different components of WM, but also serves as an intermediary between WM and LTM (Baddeley, 2012).

Other models of WM center on the focus of attention. Rather than having it be a component of a larger model as seen in Baddeley (2012), they consider WM as an attention buffer that has a storage capacity of 4 +/- 1 items (Cowan, 2001). Some researchers have also found a connection between WM and the capacity to focus one's attention while still considering both capacities to be functioning separately (Engle, 2018). Regardless, the capacity to focus attention is a particularly relevant skill in learning an L2 in an instructed setting. In fact, the underlying WM processes are useful skills in L2 learning as an adult and are used constantly in our daily lives. Following directions, reading and summarizing information, even planning a date makes use of skills such as focusing attention, remembering strings of information, and synthesizing new information with old that draw on WM resources.

1.2 Working Memory and Second Language Acquisition

From a practical standpoint, various components of WM would be relevant in a language-learning classroom. For example, take a class that is learning about comparison words in English (for example, words and phrases such as “similarly”, “neither/nor”, and “just as”). First, the teacher gives a presentation in which she introduces the vocabulary and gives examples, pointing out as she goes any particular grammar structures and punctuation that is essential to using each word or phrase. The students then go through a practice worksheet in which they compare high school with college, using words indicated to them on their worksheet. Finally, the students take turns writing their sentences on the board and the teacher leads the class through evaluating them together. The students’ WM must work with their LTM to incorporate these new words and phrase structures and build off of what they already know in English. As they go through the initial presentation, students are constantly updating the ways in which they can use the knowledge they already have; in this case, they should compare the meanings and usage of new phrases to those they already know how to use, building up new knowledge using their current baseline of grammar and vocabulary. In order to apply their new knowledge, the students must combine what they see presented on the slides and what they hear from the teacher to build a more full representation of contexts and proper usage of the different comparison words and phrases. The information must be processed in the phonological loop and visuospatial sketchpad, then combined in the episodic buffer to create the full representation. As they get feedback from the teacher and their peers on their completed sentences, they continue to update their representations on the usage of different terms.

Previous research has been able to associate different aspects of WM with a variety of functions necessary for language learning. Masoura and Gathercole (1999) found a positive correlation between foreign vocabulary learning and phonological working memory (which involves the phonological loop described by Baddeley's model). Li (2017) noted that different components of WM capacity are relevant to different areas of L2 acquisition. Li also noted a connection between the phonological loop and vocabulary learning, as well as broad connections between "complex working memory" and grammar learning and complex WM and reading comprehension. Complex WM is defined as involving both processing and storage, which would invoke the CE and the episodic buffer, as opposed to something like the phonological loop that is more concerned with simple short-term storage. Li also made a general observation in analyzing previous research that a higher WM capacity correlates with a more efficient ability to process feedback in an instructed setting (Li 2017). More generally, Serafini & Sanz (2016) found a positive relationship between WM capacity and morphosyntactic learning in beginner L2 students; Coughlin & Tremblay (2013) saw a relationship between WM capacity and the ability to recognize agreement violations in L2 students of varying proficiency; Linck & Weiss (2011) observed that WM capacity successfully predicted the degree of improvement in college students' vocabulary and grammar skills over time; and Tagarelli, Borges-Mota, & Rebuschat (2011) discovered a correlation between WM capacity and the ability to determine rules for novel syntactic patterns when actively searching for them in an artificial language.

However, although empirical links have been observed between WM capacity and the ability to learn an L2, the connection between WM training and Instructed Second Language Acquisition (ISLA) is relatively new for the field (Colflesh, Karuzis, & Rourke, 2008; Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2013). In Cognitive Science itself, there is not a substantial body of evidence demonstrating that WM can be trained, that skills acquired in WM training can be transferred, or that the effects of training WM will transfer to unrelated tasks involving WM, such as learning a second language (Jaeggi, Studer-Luethi, Buschkuehl, Su, Jonides, & Perrig, 2010).

1.3 Working Memory Training and Far Transfer

There are also those who have found a lack of evidence for transfer from WM training to other cognitive domains (Melby-Lervag, Redick, & Hulme, 2016; Sprenger, Atkins, Bolger, Harbison, Novick, Chrabaszcz, Weems, Smith, Bobb, Buntin, & Dougherty, 2013). Others have concluded that there *is* transfer to other cognitive capacities (Jaeggi et. al, 2008; Lilienthal et. al, 2013)

1.3.1 Arguments in Favor of Far Transfer

A variety of experiments have provided evidence of improvement in different skills that are seemingly unrelated to the training task; however, there is little agreement about which skills can be improved with WM training. Jaeggi et. al (2008) used the dual n-back task that has become popular as a WM training task. An n-back task, as described in section 2.3, presents stimuli rapidly and sequentially, with only one item available at a time. A single n-back task only uses one type of stimulus, usually visual; a dual n-back task uses both visual and auditory stimuli. As the series progresses, participants are asked to indicate if the item they see or hear was presented *n* items back. That is, they will be told at the beginning of the series to press a button when the item is the same as the one presented 2 items back. Jaeggi et. al trained participants in the dual n-back task for 8, 12, 17, or 19 days, and saw corresponding practice effects in n-back performance. When participants also completed an unspecified task to measure gF, their performance in that task improved with the amount of WM training participants had completed. Researchers hypothesized that gF and WM share capacity constraints, related attentional control processes, and similar neural networks, which could contribute to the observed relation between performance in the two tasks.

Lilienthal et. al (2013) found evidence of transfer to the capacity of focusing attention. This experiment also used a dual n-back task for training, and found evidence of transfer to a running digit span, which they posit is a measure of the ability to focus attention. (Curiously, although Lilienthal et. al conclude far transfer to the focus of attention, they offer evidence evidence *against* the transfer to fluid intelligence seen in Jaeggi et. al (2008) (Lilienthal et. al, 2013)).

Other studies have seen evidence of far-transfer from WM training to real-world tasks such as reading comprehension. Novick et. al (2013) looked at the role of executive function (EF), the part of WM that is likened to cognitive control; specifically, they wanted to see if WM training could help with conflict resolution, particularly with regard to language processing. They measured this capacity with the real-world ability of disambiguating temporarily ambiguous sentences (known in psycholinguistics as “garden path sentences”, as in (1) below).

(1) While Mary was mending the sock fell off her lap.

Eight different tasks were included in training. Four were programmed to challenge and hopefully improve the functioning of EF (an n-back task, a running span, a block span, and a letter-number sequencing task), and the other four came from Posit Science, a company that promoted brain training games (tasks used were from software packages *Brain Fitness Program* and *Insight*). The researchers found that participants who showed improvement in the n-back task over the course of the training regimen had more accurate comprehension of ambiguous sentences than those in the no-contact control group or those who showed no improvement in training. These participants also were faster in processing ambiguous sentences, which they measured with eye-tracking software. Similarly, Chein & Morrison (2010) observed that participants who trained with a letter span and a symmetry span task presented improvement in their reading comprehension skills. They also noted improvement in performance on the Stroop task, which measures cognitive control and inhibition. They submit that since there was clear evidence of improvement in two very different tasks that are seemingly only marginally related to the training tasks, WM training contributed to transfer to a domain-general mechanism. Specifically, Novick et. al (2013) theorize that WM training affected a mechanism related to attention that coordinates and maintains information in the face of extra processing demands.

1.3.2 Arguments Against Far Transfer

While there are several studies that argue for far-transfer effects, there are equally as many that argue against them. One of the most common arguments against evidence of far transfer from WM training, aside from several null results, is that far transfer can only be seen in tasks that share similar features or procedures with the training tasks. Sprenger et. al (2013) used a battery of tests for both training and pre- and post-testing, but found that the only evidence of transfer was in tasks that had the same stimuli or very similar aspects to one or more of the training tasks. The battery of pre- and post-testing tasks included measures of WM, inhibition, verbal reasoning, and verbal skills. Despite substantial improvement on training tasks, the only increase in scores from pre- to post-test happened for the operation span and the symmetry span. The researchers hypothesize that this is because the task of remembering serially-presented letters in the operation span was also used in three of the eight training tasks they used, and remembering the locations of serially presented stimuli was relevant to the symmetry span as well as two of the training tasks. In a further experiment within the same study, Sprenger et. al repeated a similar design with the tests and training, this time accounting for updating, resistance to interference, and visuospatial WM. However, they had different groups of participants training with different tasks, and found that the tasks that the participants improved on in the post-testing phase depended on which tasks they trained with.

Redick, Shipstead, Harrison, Hicks, Fried, Hambrick, Kane, & Engle, (2013) used an adaptive dual n-back task, an adaptive visual search group as an active control group, and a no-contact control group to measure the effects of WM training. An adaptive task becomes more challenging as the participant increases their ability, as opposed to a non-adaptive task in which the difficulty is stable and independent from the participant's performance. Their battery of pre-, mid-, and post-tests included 17 tasks measuring a combination of fluid intelligence (using logic and problem-solving), crystallized intelligence (applying previously-learned facts and knowledge), WM capacity, multitasking, and perceptual speed. Although there were clear practice effects on the training tasks, there seemed to be no transfer at all to performance on any of the 17 tasks. This differs from many other studies that argue against far-transfer, because here researchers also found no evidence of near transfer to other tasks measuring WM capacity. Specifically, there was no transfer from the n-back task used in training to either a running span or a symmetry span task. The researchers justify this by saying that although both n-back and span tasks measure some portion of WM capacity and have both been found to correlate with measures of gF, if they do not share many or any underlying processes, there may not be a reason for transfer between them (Redick et. al, 2013).

Melby-Lervag et. al (2016) conducted a meta-analysis on 87 studies on WM training published between 2002-2015. Each of the studies had a control group and cognitive pre- and post-tests. Researchers conducted analyses to mitigate effects of age, duration of training, and cognitive capacity of the participants, as well as effects from the design of the study, biases in publication, and the type of WM training program used. They were able to conclude that WM training can cause short-term improvements in verbal and visuospatial WM. However, after a few months, the training effect on verbal WM disappears, and the effect on visuospatial WM only remains for post-testing tasks that shared features with training tasks. Two of the cases analyzed noted effects of far transfer. One saw improvement in reading comprehension immediately after training, and the other saw improvement in arithmetic skills not after training but in the follow-up after several months. However, Melby-Lervag et. al had reason to doubt these results due to a pattern of decreasing scores from pre- to post-test in the control group for each study. Like many others, they conclude that although there might be near-transfer effects, "there is no evidence that working memory training convincingly produces effects that generalize to important real-world cognitive skills" (Melby-Lervag et. al, 2016, p. 523). They also note a lack of effects of personal characteristics (i.e. age or learning style) and training procedures on far transfer, but mention that these factors did have an effect on task-specific and near-transfer.

When it comes to language specifically, some studies have looked into the impact that WM training has on language processing and ability, finding a positive correlation between training and processing improvement (Novick et. al, 2013; Colflesh et. al, 2008).
Therefore, in this study, we search for evidence of transfer from one WM task that has been previously used in training this capacity, the n-back task (Jaeggi et. al, 2010), to gain in

grammatical knowledge by comparing the performance of three groups of beginner learners of Spanish from an English-speaking background during one semester of instruction.

2. Methods and Design

2.1 Participants

Participants were recruited from 4 sections of Spanish 001 at the University of Vermont. Those who consented to being part of the study but had no interest in completing any training sessions were assigned to the passive control group; they did not complete any WM training or cognitive tests, but did take the Spanish grammar exam described in section 2.3. 93 (ninety-three) students took at least the pre- or post-test exam, and 67 (sixty-seven) finished both tests. Students who were interested in completing WM training were randomly assigned to two conditions: adaptive WM training (increasing difficulty; target condition) and non-adaptive WM training (consistent difficulty; active control condition). Of the 67 students with grammar pre- and post-test scores, only 27 expressed interest in WM training; however, due to attrition, low participation in training, and one participant being far outside the average age range for this study, only 11 participants completed all of the necessary cognitive measures.

2.2 Procedure

In the first phase of the experiment, participants were given a Spanish grammar test toward the beginning of the semester to assess their baseline Spanish grammar level. Following that, they were directed to complete the online Language Background and Assumptions Questionnaire described in section 2.3. Participants then attended an individual in-person testing session in which they completed the Automated Operation Span Task (AOSpan- Unsworth, Heitz, Schrock, & Engle, 2005), which is a baseline test of WM capacity; the Flanker Task (Eriksen & Eriksen, 1974), a measure of the focus of attention and conflict resolution; the Raven's Advanced Progressive Matrices (RAPM- Raven, 1990), a measure of fluid intelligence (gF); and the N-back baseline task (Jaeggi, 2010), which measures updating capacity, a specific part of WM. All tasks are described in section 2.3 A latin-square design was used to avoid any ordering effects during testing.

Once all participants had been tested, they were given access to the online training platform with instructions to complete five training sessions per week for four weeks. Participants in the target condition completed an adaptive version of the n-back task, and participants in the active control condition completed a non-adaptive version of the same task. (For a description of adaptive versus non-adaptive tasks, see section 1.3.2 and section 2.3.) They were offered $1 (one dollar) per completed training session as compensation. The instructors of each section of Spanish 001 also gave their students extra credit at their discretion for completing training and the Spanish grammar tests. After those four weeks of training, participants had another in-person individual testing session where they performed the same four cognitive tasks again. At this time, participants also took a parallel version of the Spanish grammar test that assessed the same grammatical concepts as the initial exam.

Two of the 11 participants who completed all of the cognitive pre- and post-testing did not complete the second Spanish grammar test, and three other participants finished all cognitive measures but completed five or less training sessions, forcing the researcher to discard those data points.

2.3 Tasks and Assessment

Language Background & Assumptions Questionnaire: This questionnaire was adapted from a version used by the Spanish program as part of an assessment to measure teaching effectiveness. It asks about participants' experience with Spanish and other Romance languages, their motivation to learn Spanish and/or other foreign languages, and beliefs about the role of memory in language learning and the potential for memory training to affect how language is learned. In addition, it gauged interest and motivation to participate in the current study. It was administered online via Google Forms, prior to testing (Appendix A).

Spanish Grammar Test: This test was developed in the UVM Department of Romance Languages and Linguistics to assess the grammar proficiency of each of the basic and intermediate language courses in the Spanish program (SPAN 001-052), and the test for SPAN 001 was adapted for this study. It consists of 30 multiple-choice items that target the structures and vocabulary to be learned during the first semester of Spanish instruction at UVM. A second parallel version was created for the post-training testing (Appendices B and C). To determine that the two versions of the test were comparable, a correlation was run on the scores of both tests. The correlation coefficient was $r=0.55$, indicating that these tests were comparable in terms of their content. Upon running a paired-samples t-test on the scores, $p<0.001$, which is highly significant and indicates that the difference in scores between the first and second exams is not due to chance (which is to be expected).

Automated Operation Span Task: In this task, participants solve a simple math equation and then they are presented with a letter that they need to memorize. These equations and letters are presented in sets varying from 2 to 6 items. At the end of each set, participants need to recall the individual letters that appeared after each equation in the same order of presentation. This test measures the processing and storage capacity of WM (Appendix D).

Flanker Task: This task is a measure of the focus of attention; more specifically, it measures conflict resolution, an EF resulting from the attentional control of the CE in Baddeley's model. Participants are presented with five arrows in a horizontal line on a computer screen, and must focus on the center arrow and indicate which direction it is facing (left or right) using two keys on the keyboard. The surrounding arrows may be pointing in the same direction (congruent) or the opposite direction (incongruent) as the center arrow. The incongruent condition creates conflict, and these items typically take longer to resolve than congruent items. The associated measure is the time to respond to congruent trials subtracted from the time to respond to incongruent trials, each measured in milliseconds ($incongruent\ (ms) - congruent\ (ms)$) (Appendix E).

Raven's Advanced Progressive Matrices: This test is a visual measure of fluid intelligence. Participants are shown a geometric pattern with a piece missing, and must select the missing piece from six to eight options. Participants are given three practice items, and then eighteen test items to complete in ten minutes. The number of correctly completed diagrams is the gF score for each participant (Appendix F).

N-back Task (baseline): This task is a measure of updating capacity, which is an executive function controlled by the CE in the WM. In this task, participants are presented with a black screen, on which yellow shapes will appear for about 300ms, with 2500ms between each shape. The shapes are presented in a continuous stream, and participants are asked to indicate when the shapes were also presented *n* positions back. There are 9 trials in one testing block. This test is non-adaptive, meaning that participants' performance does not affect the value of *n*. For this test, the first three testing blocks are always 2-back, the next three are 3-back, and the final three are 4-back. This is the same test that was used for the non-adaptive training task. The score for this task and the n-back training task is calculated using $\frac{Hits - false\ alarms}{number\ of\ testing\ blocks}$, and is labeled as DV (dependent variable). Hits are the number of times the participant correctly identified that a figure was the same as *n* figures back, and false alarms are when the participant indicated that a figure was seen *n* figures back but was incorrect (Appendix G).

N-back Task (training): In this task, participants are presented with a black screen, on which yellow shapes will appear for about 300ms, with 2500ms between each shape. The shapes are presented in a continuous stream, and participants are asked to indicate when the shapes were also presented *n* positions back. In the adaptive version of this test, there are 15 trials in one testing block. If the participants complete a trial with at least 90% accuracy, the value of *n* will increase by 1. If the score is 75% or below, the value of *n* will be reduced by 1. In the non-adaptive version, there are 9 trials in each testing block. The value of *n* will increase regardless of performance as described above, with *n* reaching a maximum value of 4. The score is measured using DV as described above (Appendix H).

3. Results

The data were analyzed to look for correlations between amount and type of training, improvement in Spanish grammar score, and changes in each cognitive measure described in section 2.3. Due to attrition, only 18 of the 27 participants recruited for WM training completed any training sessions at all. Of those 18, one participant was considered an outlier from the average age range of 18-22 and therefore the researcher was forced to discard those data points due to a possible effect of age on WM capacity (Brehmer, Westerberg, & Backman, 2012). Only two of the remaining participants completed more than 5 sessions of adaptive training, and six completed ten or more sessions of non-adaptive training (as part of the active control condition), leaving a total of 8 participants who completed sufficient training sessions between pre- and post-testing. The scores for each participant after pre- and post-testing in both Spanish grammar and various cognitive capacities are shown in **Table 1**.

Table 1

Participant Number	Group	Training Sessions	SPGR T01	SPGR T02	SPGR net	AOSpan T01	AOSpan T02	AOSpan net	Flanker T01	Flanker T02	Flanker net	Nbbase T01	Nbbase T02	Nbbase net	RAPM T01	RAPM T02	RAPM net
1029	adapt	8	13	19	6	49	42	-7	79.53	75.33	4.2	0.89	2.33	1.44	12	11	-1
1027	adapt	18	15	20	5	35	49	14	150.64	96.03	54.61	-1.56	3	4.56	7	12	5
1092	non-adapt	10	19	17	-2	29	33	4	106.57	54.64	51.67	-0.78	-1.44	0.66	9	7	-2
1054	non-adapt	15	16	24	8	33	43	10	149.69	56.64	93.05	0	-1.44	-1.44	10	11	1
1034	non-adapt	16	12	13	1	35	33	-2	50.1	-41.35	91.45	1.11	2	0.89	2	2	0
1020	non-adapt	18	12	null	null	40	36	-4	47.63	101.22	-53.59	3	4.89	1.89	9	12	3
1010	non-adapt	19	9	null	null	46	42	-4	29.41	18.55	10.86	1.78	4.78	3	6	8	2
1075	non-adapt	20	10	18	8	29	28	-1	84.88	117.08	-32.2	-1.78	-2.44	0.66	10	null	null

This table contains the data from each participant in pre- and post-testing phases. Participants in the "adapt" group received adaptive training, and those in the "non-adapt" group received non-adaptive training. T01 is the pre-test, T02 is the post-test, and the net score is the difference between them. SPGR is the Spanish grammar test, and Nbbase is the n-back baseline task.

3.1 Individual-Participant Tracking

In SLA, current practice emphasizes the need to analyze data through both longitudinal studies of individuals and a snapshot of the entire group at a fixed point in time. Lowie and Verspoor (2019) claim that in a study involving human test subjects, the entire group and each individual can almost never be said to be ergodic. For a group to be "ergodic", a longitudinal study of an individual and a study of a group at one point in time would have to yield the same results upon data analysis; if the participants in this study were an ergodic group, for example, the trends observed with regard to near and far transfer would also be true of each individual participant. However, overall trends in data obtained from many people and averaged together will almost never show how each individual will change and develop over time. Therefore, group statistics cannot be applied to an individual, and individual statistics cannot represent more general data trends if the groups involved do not represent an ergodic ensemble. Many researchers have tried to measure individual differences (such as WM capacity) and use them to explain how an individual might grow and change over time, especially with respect to their learning capabilities. However, even participants with seemingly identical measures of individual differences at a given point will vary wildly in their development over time (Lowie & Verspoor, 2019).

Originally, the intention of this study was to be able to generalize results on the effect of WM training on measures of near and far transfer, with the hope that we would find evidence of far transfer to skills useful for learning a second language as an adult in an instructed setting.

However, due to the low number of participants, the data are not able to be applied to the general population as was originally intended. Therefore, in order to get a more accurate perspective on how WM training may affect both acquisition of grammar and other cognitive capacities, the analysis was carried out on the data collected from the group as a whole at the beginning and end of the study, as well as the data from each training session of eight (8) different participants.

3.2 Analysis and Correlations

Across the non-adaptive condition, there was no significant relationship between training, Spanish grammar knowledge gain, and any of the four cognitive measures used. Traditionally, ANOVAs (analyses of variance) or ANCOVAs (analyses of covariance) are used to determine the relationships between these variables; however, due to the small sample size of participants, these tests were not feasible (A. Howard, personal communication, March 25, 2019). Instead, a simple test of correlation was used. Correlations were run between Spanish grammar knowledge and each of the cognitive measures, as shown in **Table 2**. The scores of only six of the participants who completed training could be included in analysis, as two did not take the grammar post-test and therefore their net grammar score could not be compared. The correlation coefficient between the changes in Spanish grammar scores (SPGR) and the changes in the automated operations span (OSPAN) is $r=0.068$. Between SPGR and the Flanker task, $r=-0.119$. Between SPGR and the n-back baseline task (Nbbase), $r=-0.111$. Between SPGR and RAPM, $r=0.467$. A value of $p \leq 0.05$ would indicate a significant relationship between two variables; none of the correlations reported achieve that level of significance. Therefore, it is evident that the difference in grammar scores from pretest to post-test are unrelated to changes in performance on any of the cognitive tests. Particularly of note is the correlation between Spanish grammar test and the n-back baseline task. Because the n-back baseline task was the same as the training task (adaptive or non-adaptive), net performance on the n-back baseline task should reflect improvement in training over the course of the study. Since there is no correlation between these tasks, it can be concluded that there is no connection between performance in Spanish grammar and the effects of WM training in the present sample.

Among the six participants in this condition who completed between 10-20 sessions, there was a range of improvement in some of the cognitive tests as shown in **Table 1**. The measure each participant improved in and the degree of their improvement did not correspond to the number of training sessions they completed. This further supports the claim that non-adaptive training does not translate into significant improvement on WM capacity overall (Redick et. al, 2013; Sprenger et. al, 2013; Melby-Lervag et. al, 2016). The range of change in grammar gain score among these participants was from -2 to +8, which did not have a relationship with the amount of training completed. This is to be expected if in fact non-adaptive training does not affect WM capacity (Jaeggi et. al, 2010; Lilienthal et. al, 2013).

Table 2

Participant	SPGR net	AOSpan net	Flanker net	Nbbase net	Raven's net
1029	6	-7	4.2	1.44	-1
1027	5	14	54.61	4.56	5
1092	-2	4	51.67	-0.66	-2
1054	8	10	93.05	-1.44	1
1034	1	-2	91.45*	0.89	0
1020	null	-4	-53.59	1.89	3
1010	null	-4	10.86	3	2
1075	8	-1	-32.2	-0.66	null
Correlation with SPGR net	--	r=0.075	r=-0.099	r=-0.116	r=-0.367
Significance (two-tailed t-test)	--	p=0.887	p=0.851	p=0.826	p=0.475

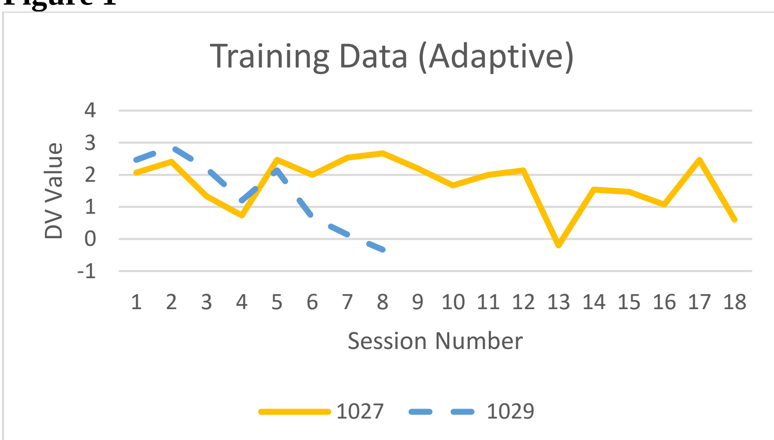
All of the net values are found by subtracting the pre-test score from the post-test score for each participant. Cells containing "null" indicate that there was no data available for that value. There were no significant correlations between SPGR net and any other net scores (at $p \leq 0.05$).
*Participant 1034's Flanker score for T02 was negative, meaning that they spent more time on the congruent than incongruent items. This is atypical for the Flanker test.

3.3 Analysis of Training Data

Throughout training, there is only one participant who showed a relatively consistent pattern of improvement over time: Participant 1010, who completed 19 sessions of non-adaptive training. Each of the other 7 participants had no identifiable trend in their training data (see **Figure 1** and **Figure 2**). This is unexpected because previous studies, even if they report no evidence of transfer effects, note a practice effect in the training task (Jaeggi et. al, 2008; Redick et. al, 2013). However, in the n-back baseline pre- and post-test, 6 of the 8 participants improved in their n-back score. The only ones whose scores did not increase were participants 1092 and 1075. There are also some participants that, even if they did not improve from the beginning to the end of training, had a consistently positive DV score. Participants 1027, 1029, 1010, and 1020 all had mostly positive scores, indicating that they had more hits than false alarms.

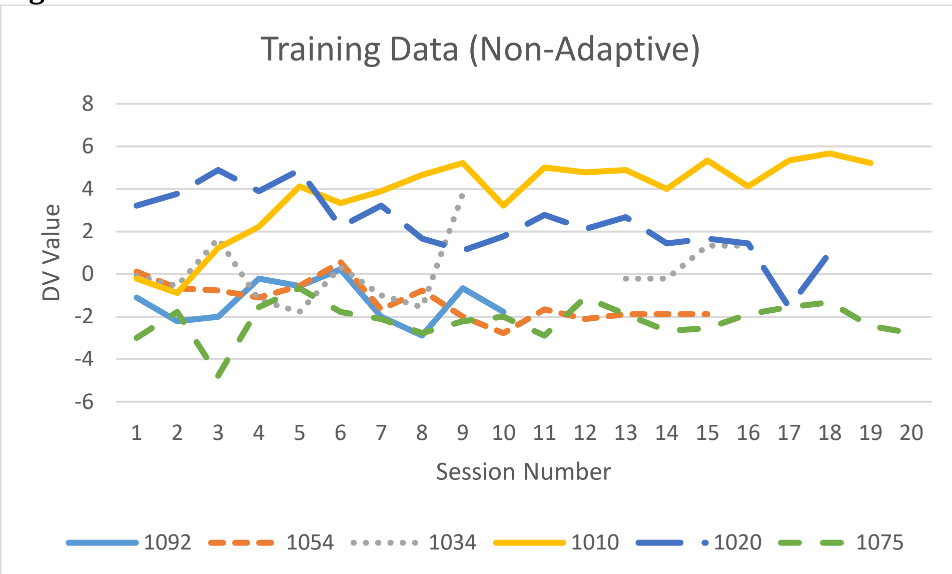
With regard to training and motivation, the data showed that there are some participants who put minimal effort into some of their training sessions. For example, during their fifth training session out of the eight they completed, Participant 1029 had 74 hits and 42 false alarms for a DV of 2.133. They reached a 3-back quickly, and completed 2 rounds of 4-back during the session. During their next session, which took place one day later, they had 28 hits and 18 false alarms for a DV of 0.667. The data collected indicates that they did not press any keys at all for several rounds of testing that day: every non-target was correct, and every target was incorrect. If left alone, the program will run until the end of each round, after which the participant must click the mouse to start the next series. This was not an isolated incident; there is evidence of this behavior from participants in both the adaptive and non-adaptive training groups.

Figure 1



DV values for each training session completed by participants in the target condition. The fluctuations in score from session to session for both participants show no clear trend.

Figure 2



DV values for each training session completed by participants in the active control condition. Only Participant 1010 (in yellow) shows a general trend, which is of gradual improvement over time. This is the expected trend due to practice effects.

3.4 Tracking Individual Participant Training and Performance

The following is a summary of the changes in cognitive and Spanish grammar scores before and after training for participants who completed non-adaptive training sessions (see **Table 3**).

Participant 1092 completed 10 training sessions, and presented an increase in their AOSpan score and Flanker test score, but a decrease in their Spanish grammar score by 2 points (6.67%).

Participant 1054, who completed 15 training sessions, improved in their AOSpan, Flanker, and RAPM scores, and increased their Spanish grammar score by 26.67% (8 points).

Participant 1034, who completed 16 training sessions, saw an increase in their performance only on the n-back baseline task, and improved their Spanish grammar score by 1 point (3.34%).

Participant 1020, who completed 18 training sessions, improved in their n-back task and RAPM performance, but no data is available on the change in their Spanish grammar score.

Participant 1010 completed 19 training sessions and improved their performance on the Flanker, n-back, and RAPM tasks. There is no data available on the change in the Spanish grammar score.

Participant 1075 completed 20 training sessions and did not improve in any of the cognitive measures in post-testing; however, they did improve their Spanish grammar score by 8 points (26.67%).

Participant 1029, who completed 8 sessions of adaptive training, improved by six points (20%) in their Spanish grammar score, and also improved in performance in the n-back baseline task and Flanker task. There is no correlation between the WM training regimen and improvement in grammar score.

Table 3

Participant	**SPGR net**	**AOSpan net**	**Flanker net**	**Nbbase net**	**Raven's net**
1029	6 (20%)↑	-7↓	4.2↑	1.44↑	-1↓
1027	5 (16.67%)↑	14↑	54.61↑	4.56↑	5↑
1092	-2 (-6.67%)↓	4↑	51.67↑	-0.66↓	-2↓
1054	8 (26.67%)↑	10↑	93.05↑	-1.44↓	1↑
1034	1 (3.34%)↑	-2↓	91.45*	0.89↑	0
1020	null	-4↓	-53.59↓	1.89↑	3↑
1010	null	-4↓	10.86↑	3↑	2↑
1075	8 (26.67%)↑	-1↓	-32.2↓	-0.66↓	null

↑ indicates an increase in score. ↓ indicates a decrease in score. *Participant 1034's Flanker score for T02 was negative, meaning that they spent more time on the congruent than incongruent items. This is atypical for the Flanker test.

3.5 Case Study

Because only one participant completed more than 8 adaptive training sessions, the lower threshold used in Jaeggi et. al (2008) to measure the effects of transfer from WM training, a detailed analysis of that participant's data was performed. Participant 1027, who completed 18 sessions of adaptive training, improved by five points (16.67%) in their Spanish grammar score and also obtained higher scores in every cognitive task between pre-test and post-test as shown in **Table 3**.

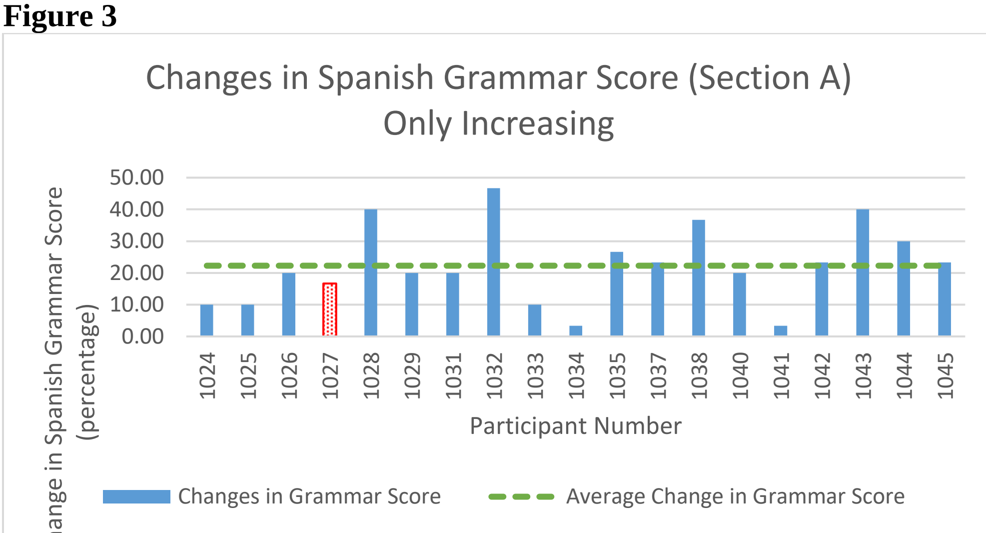
A contributing factor to Participant 1027's success in and dedication to the training regimen could be their motivation. At the time of the study, this student was in their last semester of college and needed a four-credit course to complete their requirements. In the Language Background and Assumptions Questionnaire, they indicated that they "strongly disagreed" that Spanish would be useful in communicating with their family and friends. However, they "strongly agreed" that it would be useful for their future career, and also indicated both a very strong interest in learning languages and a strong belief in the importance of memory in language learning and the potential of memory training to aid in language acquisition. In the researcher's experience, these sources of motivation can be conflicting. Learning a language for the sake of learning the language, for interest and enjoyment, usually indicates an internal drive that can sustain through the trials and tribulations of language learning. However, it can be the case that if a student has no real hope for using the language in question in real-life communication, their motivation will be much less. Learning a language for the potential of using it for future employment could be a draw to some, but could make the learning feel like an obligation to others.

The researcher contacted this participant again after training and testing in order to better understand the resulting data profile. Curiously, upon further discussion, this participant stated that they did not believe participating in memory training would actually improve their memory, which seems to contradict their earlier statement that training one's memory aids in language acquisition. Also, Participant 1027 revealed that although they believe memory is important for language learning, they do not believe they have a "good memory" despite being proficient in Mandarin and having learned it in an instructed setting. It is also worth noting that this participant did have some prior exposure to Spanish, claiming to have taken it in high school (approximately 5 years prior to taking Spanish 001 again at the time of this study), but failed their high school Spanish class. In summary, this participant was motivated to learn languages in general, but had no particular attachment to Spanish over other languages, and although they believed memory was important to language learning, did not believe in the potential for the training regimen to improve their WM capacity. The fact that Participant 1027 reported to have already taken and failed Spanish classes in the past, combined with the fact that it was a class taken to fulfill graduation requirements, could override the intrinsic motivation to learn languages in general and contribute to a lack of effort in class and on assignments.

At first, it may seem that Participant 1027's improvement in grammar can also be attributed to the effects of the WM training regimen, which is what this study hypothesized. However, upon looking at that student's improvement compared with the rest of their class, they were slightly below average in terms of mean change in grammar score. Of the 20 people in that section who completed pre- and post- Spanish grammar tests, 19 of them showed an increase in score from the first test to the second. The average increase among those students was 22.28%; participant 1027 improved by 16.67% (see **Figure 3**). Including the one student whose score decreased over the course of the semester, Participant 1027 is still below the class average in terms of improvement, although that average is slightly lower at 20.38% (see **Figure 4**). Compared to all 67 students who completed both versions of the Spanish grammar test, in both the passive and active control groups, Participant 1027 was still below the average improvement; however, this was only the case among the 58 of those students who showed some or zero improvement over the semester (average improvement = 18.79%; see **Figure 5**). Compared to all SPAN 001 students, including the 9 who scored worse at the end of the semester than at the beginning, Participant 1027 was slightly *above* average in terms of improvement (average improvement = 15.12%; see **Figure 6**).

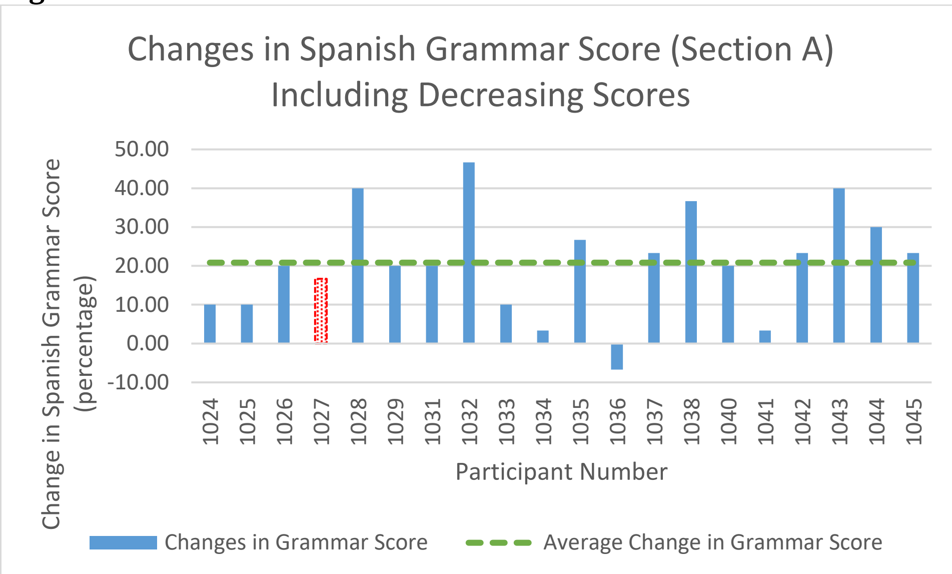
This exemplifies the fact that due to the low number of participants, we were unable to discard the possibility that improvement in Spanish grammar knowledge was motivated solely by the effective performance of the Spanish 001 instructors. We lack the data to compare Participant 1027's performance on the Spanish grammar tests to others who completed the same amount of adaptive training.

Figure 3



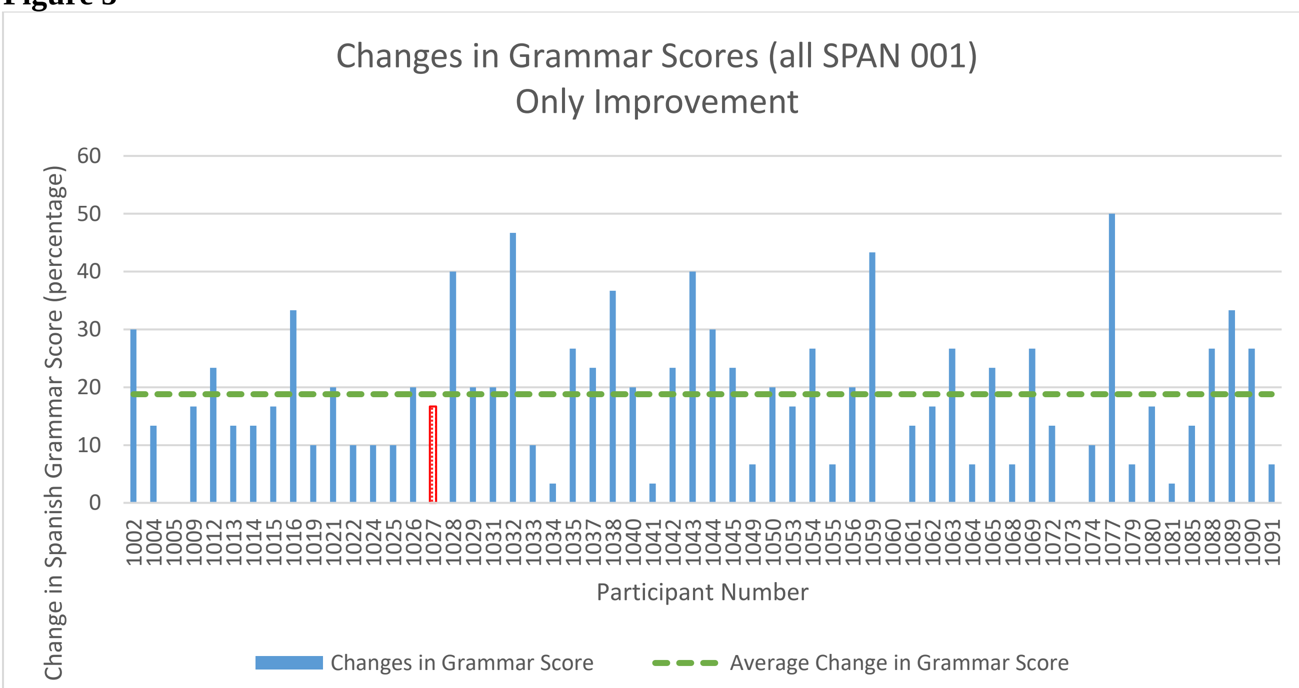
Compared to the other students in their section that improved from their first to their second grammar test, Participant 1027 is below the average of 22.28%.

Figure 4



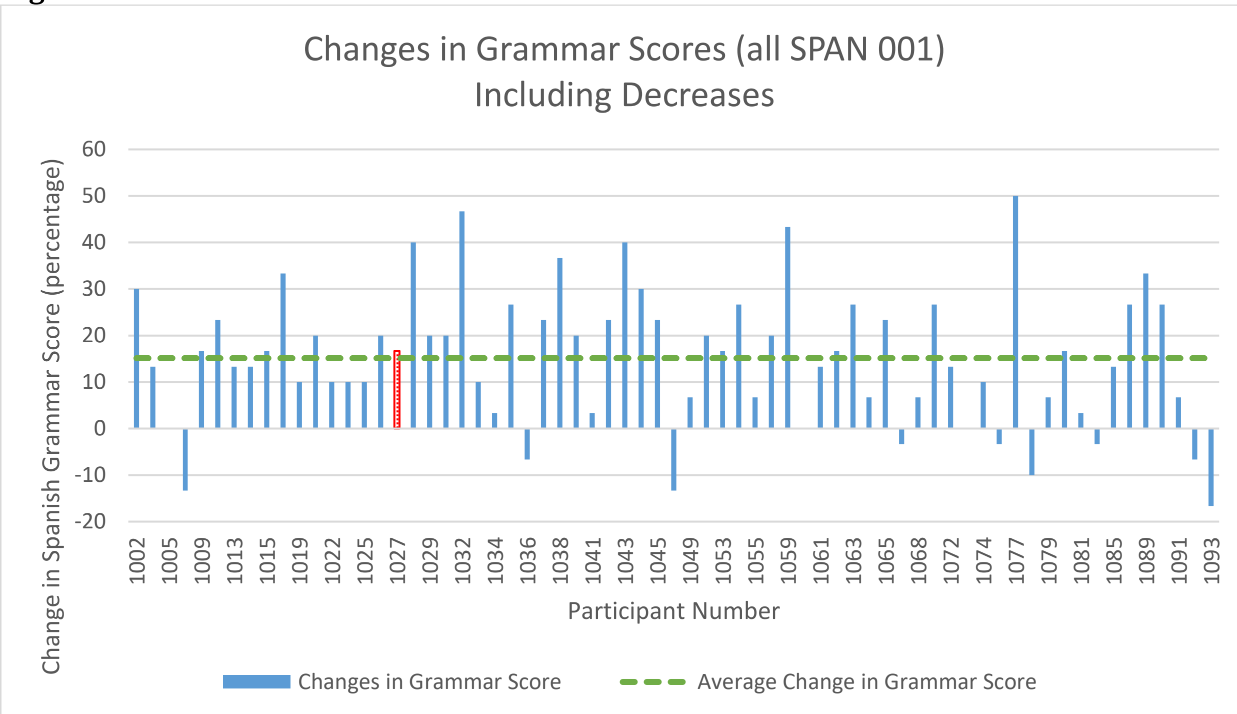
In comparison with all students in their section, including one whose Spanish grammar score decreased from the first to the second test, Participant 1027's improvement was below the average of 20.38%.

Figure 5



Among all students in SPAN 001 that increased in score between the first and second grammar tests, Participant 1027's improvement was below the average of 18.79%.

Figure 6



Including the nine students in SPAN 001 that performed worse on the second test than the first, Participant 1027's change was slightly above the group average of 15.12%.

4. Discussion

4.1 Extrinsic effects on motivation

With the methodology used in this procedure, there was a potential for a placebo effect. That is, it is expected that the active control group will show little improvement in WM capacity and less improvement than the target group on final n-back performance. However, participants in the active control group might perform better than expected because they *believe* that their training is improving their memory. A study performed by Tsai, Buschkuehl, Kamarsu, Shah, Jonides, and Jaeggi (2018) tested for a placebo effect in WM training involving an n-back task. Researchers assigned participants to one of four conditions: WM training with positive expectancy, WM training with negative expectancy, an active control group that had a positive expectancy, and an active control group with negative expectancy. The active control group trained with a task where they learned vocabulary and general facts. Groups with positive expectancy were taught about the potential for far transfer before the training started, and groups with negative expectancy were taught that training only produces near transfer or task-specific transfer. The target condition groups together outperformed the active control groups in a visual n-back task, which they trained on, and an auditory n-back task, which they did not train with. Tsai et. al also compared the performance of the WM negative expectancy group with the active control positive expectancy group on the auditory n-back task. Even though the WM training group expected no transfer effects, and the active control group did, the target condition still outperformed the active control group. Overall, they concluded that there was no evidence of a placebo effect.

Since our experiment has a similar setup to theirs, we would not expect to see much improvement in WM capacity by the participants in the active control condition due to a placebo effect in our current research project. If there had been a placebo effect in this study, participants in the active control condition would have performed consistently well in the various cognitive post-tests, regardless of the fact that their training was non-adaptive. However, because a maximum of four participants from the active control condition per test showed improvement, and the participants that did show improvement were not consistent from test to test, there cannot be a claim of a placebo effect.

Another factor that has been shown in the past to have potential effects on performance and transfer in WM training tasks is monetary compensation. A study by Katz, Jaeggi, Buschkuehl, Shah, & Jonides (2018) explores the effect of compensation on baseline, training, and transfer tasks involving WM training. They offered up to $352 for completing training; $72 for the pre-training cognitive measures, $72 for the post-training cognitive measures, and $10 for each of the 20 training sessions in between (which adds up to a total of $344). Overall, they found that compensation was positively correlated with some higher scores on baseline visuospatial WM tasks, but there was ultimately no effect on any transfer tasks. However, in their own study, the researchers question how compensation could have affected the participants' own intrinsic motivation, positing that regardless of payment, the majority of their participants had an intrinsic desire to perform well in the tasks. Compensation was offered in our own study; however, the compensation was minimal compared to Jaeggi et. al, and we would not expect a compensation effect. Compensation would increase by $1 (one dollar) for each training session completed by the participant, in addition to the extra credit offered by the participants' Spanish professors. It is possible that rather than the compensation promoting an increased desire to do well, participants were motivated simply to complete the training sessions however they could, because the $1 per training session was not worth the amount of effort that completing a session properly took (in terms of time, each session took about 20 minutes). This could happen if the participants were more extrinsically than intrinsically motivated. We speculate that this could explain the lack-of-response noted in some training sessions (as described in section 3.3).

4.2 Task-specific transfer

As shown in **Table 3**, there is evidence of task-specific transfer in that five of the eight participants in this study showed improvement in n-back performance due to training. Participants 1054, 1075, and 1092 did not improve their n-back score. In looking at their language background and assumptions questionnaire, there is nothing in these participants' self-reported answers that would indicate particularly low motivation, which could be cause for a lack of improvement in the final score.

Because four participants showed an increase in n-back score but did *not* have a trend of improvement in their training, it is possible that the environment could have had an effect on the participants' focus and motivation. Because the pre- and post-tests were administered in an office space under the supervision of a researcher, and the training sessions were to be completed at home at their own pace, unsupervised, it is possible that there was less effort put into some training sessions. There is some evidence of this in results from individual training sessions that seem unusual in the behavior of the participant, as described above in section 3.3. Like compensation, this could be an instance of extrinsic versus intrinsic motivation: in the pre- and post-testing sessions, participants knew they were being observed, which likely contributed to their extrinsic motivation to do well. However, during training sessions, any motivation to do well would have been much more intrinsic than extrinsic, leading to greater variation in performance from day to day and training session to training session.

4.3 Far transfer

Evidence for far transfer was found in a controlled testing environment for traditional cognitive tasks, but there were no effects of far transfer to second language acquisition. There was no consistent evidence of far-transfer effects from participants in the active control condition. Other studies have also concluded that the non-adaptive n-back task has minimal effects on changes or improvements to a domain-general WM capacity (Lilienthal et. al, 2013). Participant 1029 only completed eight sessions of adaptive WM training, and of those eight sessions, the last three showed evidence of lack-of-response on some or most of the series. Many researchers would argue that five complete sessions are not enough to trigger a far transfer effect (Colflesh et. al, 2008; Jaeggi et. al, 2008; Jaeggi et. al, 2010; Lilienthal et. al, 2013). However, our case study of 1027 did show evidence of far transfer to gF (as predicted by Jaeggi et. al, 2008) and focus of attention (as predicted by Lilienthal et. al, 2013). This participant also showed evidence of task-specific transfer in updating to the n-back baseline task, and of near transfer in improvement in general WM capacity.

Beyond the results of the cognitive tests themselves, Participant 1027's change in Spanish grammar score did not reflect the same magnitude of far transfer shown in their cognitive results. Had there been evidence of transfer to skills useful to second language acquisition, we would have expected to see Participant 1027 rise above their peers in their Spanish class due to the usefulness of the various cognitive capacities tested as outlined in section 1.2. However, Participant 1027's degree of improvement was below average for their section of Spanish 001. This is in keeping with findings from various authors (Redick et. al, 2013; Sprenger et. al, 2013; Melby-Lervag et. al, 2016) who argue that transfer does not occur unless the training task shares stimuli or methods with the task used to measure transfer. Since in this case the task used to measure transfer is a Spanish grammar test, which differs greatly from an n-back task in terms of stimuli and procedures, our findings replicate other conclusions that show a lack of transfer from WM training to everyday tasks involving skills or other cognitive processes that might have been improved by WM training.

5. Conclusion

5.1 Summary

In conclusion, this study adds to the body of evidence that shows that although WM training with a specific task might have far-transfer effects in a formal, controlled testing environment, it does not transfer to second language acquisition *in this context*. There is evidence of task-specific transfer, practice effects which are commonly seen throughout the WM training literature (Jaeggi et. al, 2008). However, there was no consistent pattern for most participants throughout the course of their training sessions. Although participants in both the target and active control conditions saw improvement in various cognitive capacities, the skills in which they improved were not consistent between participants, and did not correlate with the amount of training sessions completed. There is a clear link, however, between adaptive WM training and improvement in domain-general cognitive capacities, as seen in Participant 1027's performance. It is essential to note, however, that our results only assess WM training with an updating task, and the effects of that training on grammar knowledge. The lack of transfer found herein does not exclude the possibility of transfer from other WM capacities, trained using different tasks, to other skills relevant to language acquisition.

5.2 Future Steps

The most significant detriment to this study was the initial lack of participants combined with high rates of attrition throughout the course of the experiment. With more participants, results could potentially be generalized to the broader population of L2 learners; as it is, the data gathered in this study are highly dependent on individual motivations and day-to-day performance. In addition to the lack of participants, it would be useful to have more contact with participants throughout the training period. This could happen in a number of ways. Participants could come in for in-person training sessions, to be monitored by a researcher. This would mitigate the lack-of-response effects seen by several participants, and having a researcher present would likely motivate the participants to perform well. In-person training sessions would also aid in moderating the time of day that participants performed training sessions. It might also be useful for students to complete all of the training either during school break or while classes were in session (but not both), as this could affect their motivation. If having in-person training sessions is not feasible due to time, space, or other constraints, and an online training paradigm is still necessary, then researchers should check up on participants more frequently and with more attention to encourage those who fall behind. It is also possible that higher compensation could increase the participants' motivation to do well, as seen in Katz et. al (2018).

As mentioned in section 5.1, the lack of evidence of transfer to SLA in this study does not rule out the possibility for other kinds of transfer between different tasks and skills involving WM. For example, as mentioned in section 1.2, both Li (2017) and Masoura and Gathercole (1999) found a connection between the phonological loop and vocabulary learning. Neither the phonological loop nor vocabulary learning was covered by the scope of the current study. It is certainly possible that WM training that improves the efficiency of auditory processing and storage could have far transfer effects to speed and retention in acquiring new vocabulary.

References

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology, 63,* 1-29.

Brehmer, Y., Westerberg, H., & Backman, L. (2012). Working-memory training in younger and older adults: training gains, transfer, and maintenance. *Frontiers in Human Neuroscience, 6,* 63.

Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review, 17*(2), 193-199.

Colflesh, G., Karuzis, V., & Rourke, P. O. (2008). Effects of Working Memory Training on L2 Proficiency and Working Memory Capacity, (1) 289-294.

Coughlin, C. E. & Tremblay, A. (2013). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics*, *34,* 615-646.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114.

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science, 13*(2), 190-193.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon identification of a target letter in a nonsearch task. *Perception and Psychophysic*, *16,* 143-149.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008) Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*(19), 6829-6833.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning – implications for training and transfer. *Intelligence*, *38,* 625-635.

Li, S. (2017). Cognitive Differences and ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (396-417). New York: Routledge.

Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, & J., Hale, S. (2013). Dual *n*-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review, 20,* 135-141.

Linck, J.A., & Weiss, D.J. (2011). Working Memory Predicts the Acquisition of Explicit L2 Knowledge. In C. Sanz & R. P. Leow (Eds.), *Implicit and Explicit Language Learning: Conditions, Processes, and Knowledge in SLA and Bilingualism* (101-113). Washington, D.C.: Georgetown University Press.

Lowie, W. M. & Verspoor, M. H. (2019). *Language Learning*, *69*(1), 184-206.

Masoura, E. V. & Gathercole, S. E. (1999). Phonological Short-term Memory and Foreign Language Learning. *International Journal of Psychology, 34*(5/6), 383-388.

Melby-Lervag, M., Redick, T. S., & Hulme, C. (2016). Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measure of "Far Transfer": Evidence From a Meta-Analytic Review. *Perspectives on Psychological Science, 11*(4), 512-534.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2013). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience, 29*(2), 186-217.

Raven, J. C. (1990). *Advanced Progressive Matrices. Sets I, II.* Oxford: Oxford University Press.

Redick, T.S., Shipstead, Z., Harrison, T.L., Hicks, K.L., Fried, D.E., Hambrick, D.Z., Kane, M.J., & Engle, R.W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. Journal of Experimental Psychology: General, 142(2), 359- 379.

Serafini, E. J. & Sanz, C. (2016). Evidence for the Decreasing Impact of Cognitive Ability on Second Language Development and Proficiency Increases. *Studies in Second Language Acquisition, 38,* 607-646.

Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J. S., Weems, S. A., Smith, V., Bobb, S., Bunting, M. F., Dougherty, M. R. (2013). Training working memory: Limits of transfer. *Intelligence, 41*(5), 638-663.

Tagarelli, K.M., Borges-Mota, M., & Rebuschat, P. (2011). The Role of Working Memory in Implicit and Explicit Language Learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*(33), 2061-2066.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498-505.

Appendices

APPENDIX A
Language Background and Assumptions Questionnaire

Participant Number: …………………………….
95# : ……………………………
Placement test score: ………….

Gender: ……………………………
Age: ………………………….
Ethnicity: ……………………………..

Why are you studying Spanish at UVM?
………
………
…………………………

Are you planning to become a Spanish major? YES NO

Are you planning to become a Spanish minor? YES NO

What is your current major? ………………………………………….

What is your current minor? ………………………………………….

You are a ….1st year .… sophomore …. junior …. senior …. other student
….Other (specify): ………………………………………..

Have you ever traveled to a Spanish-speaking country? YES NO

Which Spanish-speaking country(ies) did you travel to??
……………………………………………………………………………………………
How long were you in each country?
……………………………………………………………………………………..

What other languages do you know (please list them below)?
………
………
…………………………

How would you rate your fluency in language 1?
Native speaker … Advanced …. Intermediate ….
Beginner ….

How would you rate your fluency in language 2?

Native speaker … Advanced …. Intermediate ….
Beginner ….

How would you rate your fluency in language 3?
Native speaker … Advanced …. Intermediate ….
Beginner ….

By learning Spanish, I will be better able to communicate with my friends and/or family.
Strongly Agree 1 2 3 4 5 Strongly Disagree

I believe that by training your memory, you can learn a language more easily.
Strongly Agree 1 2 3 4 5 Strongly Disagree

Learning Spanish will be useful for my future career.
Strongly Agree 1 2 3 4 5 Strongly Disagree

I believe that a good memory is important for learning a new language.
Strongly Agree 1 2 3 4 5 Strongly Disagree

I am generally interested in learning new languages.
Strongly Agree 1 2 3 4 5 Strongly Disagree

I would be interested in participating in a study that examines the effects of memory training on the ability to learn a new language.
Strongly Agree 1 2 3 4 5 Strongly Disagree

APPENDIX B
Spanish Grammar Test: Version 1

<u>SPAN001 - EXAMEN DE GRAMATICA</u>

1. A Juana no ________________ gustan las películas de ciencia ficción.
   A. le
   B. se
   C. la
   D. lo
2. A= ¿De dónde es usted? B= ___________ de Perú.
   A. Estoy
   B. Está
   C. Ser
   D. Soy
3. A= Mucho gusto. B= El gusto es ____________
   A. ello
   B. mío
   C. más
   D. me

4. Patricia viene de Ecuador. Pedro y Elena ____________ de Guatemala.
   A. seis
   B. sois
   C. son
   D. eres
5. 1:15pm Es la una y ____________
   A. diez
   B. diez y cinco
   C. quince
   D. quinto
6. En los EE.UU. __________ 9.2% de hispanos que son de Puerto Rico.
   A. hay
   B. ahí
   C. es
   D. ser
7. Me encanta ______________ panqueques para mis niños.
   A. hago
   B. hay
   C. hacer
   D. hace
8. _____________ televisión en mi computadora por la noche.
   A. Prendo
   B. Escucho
   C. Miro
   D. Mío
9. ¿Te _______________ la sociología?
   A. gusta
   B. gustas
   C. gusto
   D. gustan
10. El teléfono __________ encima del escritorio.
   A. es
   B. está
   C. estás
   D. eres
11. Después del receso de primavera, ____________ a la universidad en autobús.
   A. olvido
   B. juego
   C. regreso
   D. llevo
12. Nosotras _____________ mochilas del mismo color.
   A. tenéis
   B. tengan
   C. tenemos
   D. tengáis
13. Los domingos __________ almorzar con mi madre en su casa.

A. puedo
B. pudo
C. podo
D. puedes

14. Noelia ___________ alojarse en una pensión cuando viaja a Madrid.
A. siempre
B. pareces
C. prefiere
D. nunca

15. Pedro está avergonzado porque su habitación ___________ un desastre.
A. es
B. está
C. estar
D. estás

Completa el siguiente texto con la mejor opción para cada espacio en blanco.

La rutina de Patricia

¡Hola! me llamo Patricia y vivo en Puerto Vallarta, México. Quiero contarte cómo es un día típico en mi vida. Por la mañana (16) ………. café con mis padres y juntos (17) ………. las noticias por la radio. A las siete y media, (18) ………. de mi casa y (19) ………. el autobús. Me (20) ………. llegar temprano a la universidad porque siempre (21) ………. a mis amigos en la cafetería. Tomamos café y (22) ………. lo que vamos a (23) ………. cada día. A las ocho y quince, mi amiga Sandra y yo (24) ………. al laboratorio de lenguas. (25) ………. clase de francés (26) ………. a las ocho y media. ¡Me (27) ………. el francés! A las doce y media (28) ………. en la cafetería nuevamente con mis amigos. Después, (29) ………. mi tarea en la biblioteca. Por las tardes, mis amigos (30) ………. a sus casas, pero yo juego al vóleibol con mi amigo Tomás.

16. A. cocino	B. bebo	C. temo	D. tomas
17. A. escuchamos	B. miramos	C. vemos	D. leemos
18. A. entro	B. limpio	C. salgo	D. voy
19. A. bebo	B. toma	C. conduce	D. tomo
20. A. molesta	B. gusta	C. enfada	D. aburre
21. A. encontro	B. encuentro	C. encontrar	D. encontramos
22. A. planeamos	B. planamos	C. volvemos	D. volvamos
23. A. hago	B. hacemos	C. haces	D. hacer
24. A. caminemos	B. caminos	C. vamos	D. vemos
25. A. Aquella	B. La	C. Mía	D. Porque
26. A. comenza	B. comiensas	C. empieza	D. comienzo
27. A. fascina	B. disgustas	C. encantan	D. fastidio
28. A. soy	B. somos	C. son	D. estoy
29. A. hago	B. escribe	C. resuelva	D. entrego
30. A. volven	B. volvemos	C. vuelven	D. regreso

APPENDIX C
Spanish Grammar Test: Version 2

SPAN001 - EXAMEN DE GRAMATICA (2)

1. Antes del receso del invierno, ____________ mis libros de texto a la librería.
A. olvido
B. devuelvo
C. regreso
D. llevo
2. Me encanta ______________ dibujos para mi clase de arte.
A. hacer
B. hago
C. hace
D. hay
3. En el verano, _____________ los pájaros en el parque.
A. llueve
B. camino
C. mío
D. miro
4. En Chile, _________ 32 universidades en la ciudad capital.
A. ahí
B. es
C. ser
D. hay
5. El lápiz __________ debajo de la mesa.
A. eres
B. es
C. está
D. estás
6. ¿Te _______________ tus clases?
A. gusta
B. gustas
C. gusto
D. gustan
7. Dioni tiene miedo porque afuera ____________ oscuro.
A. está
B. es
C. estar
D. estás
8. Los lunes __________ estudiar con mis amigos en la biblioteca.
A. podo
B. pudo
C. puedes
D. puedo
9. Cris ____________ viajar en tren cuando va de viaje en Santiago.

A. siempre
B. pareces
C. nunca
D. prefiere
10. A nosotros no _________________ gusta el chocolate con leche.
A. lo
B. nos
C. les
D. os
11. Nosotras _____________ libros del mismo autor.
A. tenéis
B. tengan
C. tenemos
D. tengáis
12. A= ¿De dónde es tu abuela? B= ____________ de Perú.
A. Está
B. Eres
C. Es
D. Ser
13. A= ¿Es esta tu bicicleta?. B= Sí, es _____________
A. de mí
B. me
C. yo
D. mía
14. Marisol viene de Venezuela. Hilario e Isabel ____________ de Uruguay.
A. seis
B. son
C. eres
D. sois
15. 12:47pm Son las doce y ____________
A. cuarenta
B. cuarenta y siete
C. cincuenta y siete
D. veintisiete

Completa el siguiente texto con la mejor opción para cada espacio en blanco.

<u>La rutina de Ofelia</u>
¡Hola! me llamo Ofelia y vivo en La Paz, México. Ahora, estoy en la escuela secundaria y quiero contarte cómo es un día típico allí. (16) ……… a la escuela a las 7:45 de la mañana, porque clase (17) ……….a las 8. Con mi amiga Camila, (18) ……… de nuestra tarea de matemáticas y (19)………… las respuestas (no lo digas al profesor). Siempre (20) …….. la clase de inglés primero, y me (21) ……….. leer las novelas. Después de inglés, Camila y yo nos (22) …… para (23) …….. a matemáticas. A las doce y media, yo (24) ……….. almuerzo de la cafetería, pero mis amigos Kevin y Pepo (25) ………. el almuerzo de casa. Por qué (26) ……. próxima clase

(27).......... lejos de la cafetería, ellos (28)......... primero. Más tarde, (29) ……… a mis clases de biología y estudios sociales, y así (30) ……… el día escolar.

16. A. Llego	B. Para	C. Pinto	D. Ir
17. A. termina	B. empieza	C. entiende	D. esfuerza
18. A. caminamos	B. hablamos	C. corremos	D. bebemos
19. A. borramos	B.llevamos	C. comimos	D. compartimos
20. A. regreso	B. escribo	C. tengo	D. bailo
21. A. encantas	B. encanto	C. encantamos	D. encanta
22. A. juntamos	B. juntemos	C. juntan	D. junto
23. A. voy	B. ir	C. va	D. van
24. A.compró	B. compro	C. compran	D. compras
25. A. traen	B. sonrien	C. piensan	D. compran
26. A. le	B. tú	C. su	D. tu
27. A. soy	B. es	C. estoy	D. está
28. A. salen	B. salgan	C. salgo	D. salo
29. A. vas	B. voy	C. ir	D. entrar
30. A. termine	B. terminas	C. terminan	D. termina

APPENDIX D
Automated Operation Span Task

Sample screens

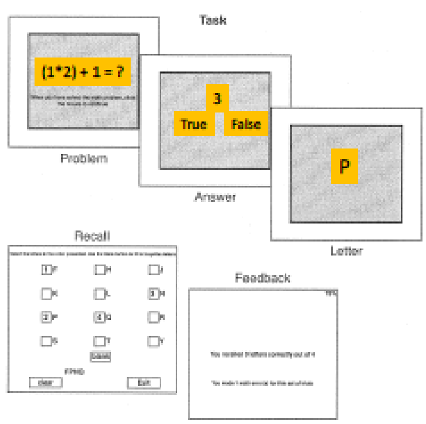
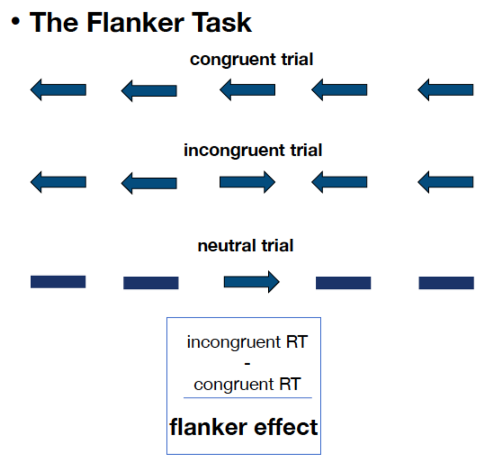


Figure 1. Illustration of the automated operation span task. In the task, first a math operation is presented. After it is solved, participants click the mouse and a digit is presented, which is judged to be either the correct or incorrect answer to the math operation. This is followed by a letter for 800 msec. For recall, the correct letters from the current set are selected in the correct order. After recall, feedback is presented for 2,000 msec.

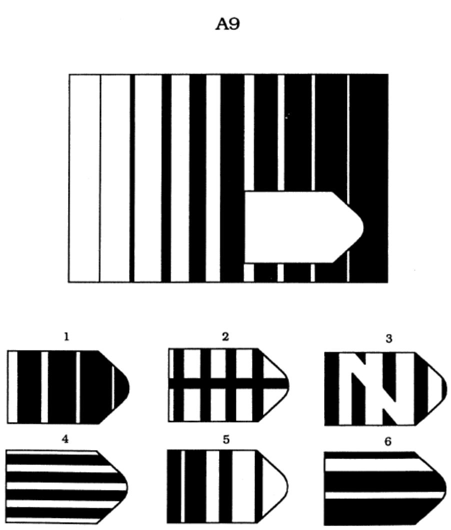
APPENDIX E
Flanker Task

Example of stimuli used in this test



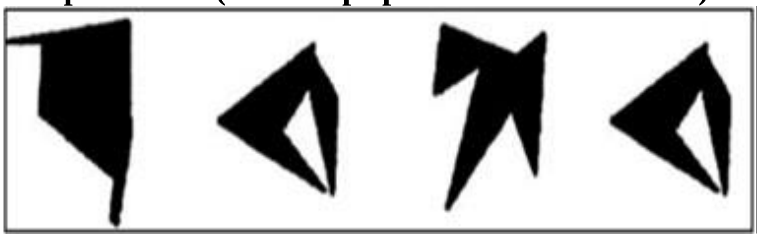
APPENDIX F
Raven's Advanced Progressive Matrices

Sample Test Item



APPENDIX G
N-Back Task (Baseline)

Sample stimuli (each shape presented one at a time)



APPENDIX H
N-Back Task (Training)

Sample stimuli (each shape presented one at a time)