EXAMINING THE BILINGUAL ADVANTAGE IN VISUOSPATIAL EXECUTIVE FUNCTION TASKS FOR REGULAR USE BILINGUALS

JESSICA A. JENSEN

Bachelor of Arts in Psychology

Kent State University

June 2006

submitted in partial fulfillment of requirements for the degree

MASTER OF ARTS

at the

CLEVELAND STATE UNIVERSITY

December 2020

We hereby approve this thesis for

JESSICA JENSEN

Candidate for the Master of Arts degree

For the Department of Psychology

And

CLEVELAND STATE UNIVERSITY'S

College of Graduate Studies by

Thesis Chairperson, Conor T. McLennan

Department of Psychology / December 10, 2020

Methodologist and Committee Member, Kenneth Vail

Department of Psychology / December 10, 2020

Committee Member, Andrew Slifkin

Department of Psychology / December 9, 2020

Committee Member, Albert Smith

Department of Psychology / December 10, 2020

Student's Date of Defense: December 8, 2020

EXAMINING THE BILINGUAL ADVANTAGE IN VISUOSPATIAL EXECUTIVE FUNCTION TASKS FOR REGULAR USE BILINGUALS

JESSICA JENSEN

ABSTRACT

We tested the hypothesis that young adult bilinguals show a significant advantage relative to young adult monolinguals in two visuospatial executive function (EF) tasks, the Simon task and the Corsi task. The focus was on bilinguals who reported being exposed to all of their languages on a regular basis (i.e., used no language more than 60% of the time). Regular language use has been identified in past theoretical positions as an important mechanism contributing to a bilingual advantage. In many previous studies of bilingualism and EF, which have produced conflicting results, researchers have included verbal components, as opposed to visuospatial tasks of working memory. Consequently, inconsistent results across previous studies could be due, at least in part, to differences in verbal abilities between monolinguals and bilinguals. Tests that use verbal components to examine EF may be probing verbal abilities, rather than EF. Visuospatial tasks, including the Simon (Bialystok, 1999) and Corsi (Berch et al., 1998) tasks, provided an opportunity to measure EF performance without the possible confounding factor of verbal ability. The data indicate patterns of a bilingual advantage in conditions that place heavier demands on EF, although not statistically significant. Results based on comparisons of bilinguals and monolinguals are tentative until a larger sample size of bilinguals is obtained.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The hypothesis of a bilingual advantage, better performance by bilingual individuals in certain executive function (EF) tasks compared to their monolingual peers (Bialystok et al., 2004), has been persistent, despite controversy and contradicting evidence. Some critics point out that only a small percentage of recent studies have obtained evidence of a bilingual advantage (Moreau et al., 2019), while other critics hold that the situations in which the advantage has emerged are so specific that it is not worth considering the implications of such an advantage in daily life outside the laboratory (Paap et al., 2015; Paap & Greenberg, 2013). In the current study, which is a response to these two particular types of criticisms, we investigated why only a portion of studies have obtained empirical support for a bilingual advantage while other researchers have been unable to obtain evidence in support of a bilingual advantage. To do so, we examined whether the advantage exists only in specific situations, and if such conditions are likely to exist outside the lab, in which case the advantage would have potentially meaningful implications in daily life.

In this Introduction section, supporting evidence for a bilingual advantage being driven by executive function will be discussed. This evidence spans from research in social development to cognitive processes, providing context for the generalizability of a bilingual advantage in EF. Discussing findings in social development research as well as cognitive processes also allows us to consider a greater picture of the mechanisms driving the advantage. We will then discuss the models that have led to current theory, followed by a discussion of the variability of the bilingual experience and how this variability must be considered in current models. Studies that show support for a bilingual advantage in EF, as well as critiques of the theory, are discussed. Finally, the design and theory of the current study are discussed.

There has been evidence that bilingual children have an advantage over monolingual peers in social skills, such as theory of mind, which is the ability to perceive how others think and feel. In false-belief (FB) tasks, participants are asked to think about the perception of others through changed location tasks or unexpected contents tasks. The changed location tasks involve a story about a character who places an object in a hidden place and then leaves. Another character then changes the location of the object, and the original character returns to retrieve their object. The child (participant) is asked where the character will look. In the unexpected contents task, participants are shown a box with desired contents, such as Smarties candies. A secret compartment with less desirable contents, such as pencils, is deployed when the content is shown to the participant a second time. The participant now sees that instead of Smarties candies, the box contains pencils. Participants are asked what a character thinks is inside the box, when the character has only seen the original candy contents and not the changed pencil contents. In variations of this task, children are asked whether or not the character will be happy or sad upon opening the box (Gordon, 2015). Bilingual children have been found to predict the character’s reactions more often than their monolingual peers (Gordon, 2015), and bilingual children seem to be able to pass the tasks at a younger age than monolinguals (Rubio-Fernandez, 2016.) Although the mechanism underlying this advantage is debated, it is generally accepted that these advantages stem from superior EF development. Rubio-Fernandez (2016) suggests that although popular belief is that the advantage stems from better performance in inhibition, the advantage actually results from enhanced attention. Benson and her team (2013) examined the role that individual differences play in developing EF to provide an advantage in theory of mind tasks. These researchers found that individual differences in experience led to EF learning, which better prepared these functions to negotiate the demands of the tasks. The language experiences of bilinguals and monolinguals have been shown to affect the outcome of FB tasks, providing further support for the hypothesis that individual differences are driving EF development. Gordon (2016) studied the role of language proficiency in FB tasks and found a relationship between high proficiency across languages, both in monolingual and bilingual groups, and better performance in FB tasks. However, the researchers of this study could not identify the mechanism, such as EF, that was responsible for better performance. Gordon’s conclusions suggest that the advantages seen in many instances may be an individual language experience advantage, as opposed to a direct bilingual advantage.

In an early study of a bilingual advantage, Bialystok (1988) compared participants with varying levels of bilingualism on aspects of linguistic awareness. Three groups of children, monolingual English speaking, partially French-English bilingual, and fully French-English bilingual were asked to complete metalinguistic tasks, such as correcting syntax errors, imagining that object labels were swapped and then having to describe characteristics of the new labels, and defining abstract meanings such as “what is a word.” The results demonstrated that fully bilingual children performed the best at these tasks and monolingual children had the lowest performance. Bialystok attributed these differences in performance to bilingualism advancing the development of executive control over language. Bialystok makes note that further investigation of bilingual advantages must consider the participants’ individual language experiences before making conclusions.

The population of bilingual individuals is so diverse that the condition of being bilingual for one participant is not the same as the condition of being bilingual for another participant. When a study includes participants with varying language experiences in one single group of bilinguals, the effect of the individual language experiences could be a confounding factor (Bialystok, 1988). Variables such as age and method of acquiring a second language (Luk et al., 2011), language proficiency (Gordon, 2016), and socioeconomic status (Czapka et al., 2019) have all been suggested to influence a bilingual advantage. One variable that stands out is that of the frequency with which a bilingual uses his or her two languages. Past studies have included measures of language use as a control, but few studies have actually examined the role of language use in the outcomes of cognitive performance (Incera & McLennan, 2018).

Previous theories derived an explanation for the bilingual advantage in EF as an exercise effect, or that certain functions of the brain get stronger when exercised or used more often. The Inhibitory Control (IC) Theory (Green, 1998) purports that in order to control their languages, bilinguals use EFs, such as working memory, that maintain information for two words rather than one, inhibition for suppressing the inappropriate language, and switching for understanding which language rules the bilinguals must operate under in any given social situation. Under such regular use, EF might be expected to strengthen, and thereby lead to better performance. Importantly, this improved performance could extend outside the domain of linguistic processing.

Further testing of the functions involved in juggling multiple languages has brought about a slightly different theory, the Adaptive Control Hypothesis (Abutalebi & Green, 2007). According to this theory, executive control adapts to the demands of the situation in which a talker is engaging. In a single language situation, for example, where speaking in one language is appropriate, there is greater demand placed on the EF of inhibition to suppress the other, inappropriate, language. Because the EF has adapted to meet demands, EF is now engaged and “primed” for any other inhibition tasks that the speaker might encounter during this situation. One might then wonder, if EF is adapting to environmental cues, whether there are individual differences that cause EFs to meet demands within the brain. The process may be likened to pulling out of a driveway in a car. If an individual gets in a car that is already running, less time is needed to shift gears and begin to move the vehicle. If a different individual enters a car and has to first put the keys in the ignition and start the engine, slightly more time is required to engage the gears and move the vehicle. This is an overly simplistic analogy, wherein the cars represent cognitive functions; however, one can imagine that processing information may be faster when the brain is already “primed” to process this information with specific mechanisms. Under the Adaptive Control Hypothesis, certain language experiences in bilingualism may allow individuals to keep their EF engines running, reducing the time necessary for processing in some situations.

A bilingual advantage has been examined in working memory (WM) on several occasions, with varying results. Ratiu and Azuma (2014) examined WM differences between bilingual and monolingual college students. Their bilingual group consisted of Spanish-English speakers who learned Spanish as a first language (L1) and English as a second language (L2). Using a series of span tasks, the researchers found that bilingualism was not a predictor of performance on WM tasks. However, the participants that were in the bilingual group reported using English, L2, much more than Spanish, L1, on a daily basis. While the lack of advantage in WM tasks may be attributed to the age of the participants, or to a lack of interaction between bilingualism and WM features, the lack of advantage could also be attributed to disproportionate rates of use between L1 and L2 (Incera & McLennan, 2018).

The lack of a bilingual advantage could also be due to the nature of the tests used to measure WM. Blom, Kuntay, Messer, Verhagen, and Leseman (2014) found that when the tasks controlled for vocabulary and socio-economic status, a bilingual advantage in visuospatial tasks such as a dot matrix and "odd one out" puzzles, as well as verbal tasks, such as backward digit recall, emerged by age six (Blom et al., 2014). These researchers used a backward digit recall as a verbal task because backward digit recall requires more executive control than the often-used forward digit recall. The bilingual advantage was measured to be the greater in the backward recall task, suggesting that the advantage is the result of a stronger executive control in tasks that require both processing and storage. Similarly, another study, by Morales, Calvo, and Bialystok (2013), showed a trend toward a larger advantage as tasks required more executive control. Using a Simon task and a visuospatial span task at varying levels of difficulty across ages 5-7, the researchers found that bilingual children held the greatest advantage at the most difficult level of the task. In the simpler versions of the task, the younger bilingual participants performed as well as the monolingual older children (Morales et al., 2013).

Warmington, Kandru-Pothineni, and Hitch (2018) wanted to strengthen the replicability of bilingual studies by providing a clean definition of bilingualism and by using a group of participants who shared a common native and second language. The researchers found that with a well-defined parameter of bilingualism, a bilingual advantage could be seen in WM tasks. Warmington et al. suggest that the age at which L2 is acquired can determine the effects of a bilingual advantage, showing that individuals with a lifelong use of two languages outperform monolinguals.

The topic of age in relation to the acquisition of L2 has also been considered. In work by Luk, Desa, and Bialystok (2011), the age that bilingualism was achieved had an effect on performance in a flanker task. Their findings suggest that individuals achieving bilingualism after the age of 10 are outperformed by individuals who achieved bilingualism earlier in childhood. While there is no such direct evidence that age affects the bilingual advantage on visuospatial WM tasks, the work of Warmington and Luk's teams demonstrates the necessity of considering participants' language experience.

The regular usage of two (or more) languages may be more strongly associated with a bilingual advantage than acquiring an L2 at an early age. In 2018, Incera and McLennan investigated bilingualism and age as continuous variables. These researchers found that when individuals were "equal-use" bilinguals, or individuals who used both

L1 and L2 approximately equally often in their daily routines, there was a stronger bilingual advantage across middle aged and aging adults than those who used their languages less regularly. Incera and M[c]Lennan, however, did not focus on young adults. Further, the researchers tested certain aspects of EF by using a Stroop task, but not a visuospatial WM task. Bialystok, Craik, Klein, and Viswanatha (2004) conducted a study on middle aged and older adults using a Simon task to measure WM performance. The results of their study supported a bilingual advantage in WM for adults, and also indicated that lifelong bilingualism may reduce age-related declines in inhibition. However, young adults were not included in their study. Alloway, Gathercole, and Pickering (2006) conducted a study to determine if visuospatial and verbal components of WM are mutually dependent, or work as separate entities within WM, and whether or not these components change as a child matures (Alloway et al., 2006). Using visuospatial span tasks, odd man out, and other WM tasks, the researchers concluded that verbal and visuospatial processing are two separate domains within WM, and that the structures remained constant throughout development.

Despite the supporting evidence discussed above, the idea of a bilingual advantage has had critics. Paap and his colleagues (2015) attempted to design a study that built upon the idea of an advantage in WM and inhibition. After their study demonstrates that the methodology failed to result in the expected effect, Paap et al. concluded that the situations in which a bilingual advantage exists in EF are so specific and weak that an effect in daily life outside the research laboratory is implausible. These critics further claim that the observed effects may not be a result of bilingualism, but instead may be due to alternative confounds, such as socioeconomic status or age.

In the current study, we focused on young adults' performance on two visuospatial tasks. The Corsi task (Berch et al., 1998) is a visuospatial WM task in which participants are shown a sequence of colored squares on a computer monitor and asked to repeat the sequence by clicking on the squares with a computer mouse. In one version, participants are instructed to recall the span in the order in which the squares were shown, and in another version, participants are instructed to recall the span in the reverse order, which requires more executive control. In the Corsi task, we examined the effects of a bilingual advantage on WM; however, the widely accepted model of EF is one in which many functions operate together (Miyake et al., 2000). Comprised of several processes that work in unity, EF would see a domino effect with demands on one process resulting in demands upon others. Therefore, looking beyond tasks that examine only one process, such as WM, and including tasks that examine many processes working together, was essential.

We examined inhibition, WM, and rule switching as individuals were instructed to remember a set of rules, switch between sets of rules, and ignore conflicting information to complete the Simon task, as described by Bialystok et al (2004). In one set of trials, participants were instructed to remember the corresponding keys to two colors (one color per key) and respond while ignoring the position of the squares relative to the position of the response key. In another set of trials, individuals were instructed to remember the corresponding keys to four colors (two colors per key) and respond while ignoring the position of the squares relative to the position of the response key.

To measure language use, participants completed the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007). This questionnaire provided insight to participants' language experiences, including how often they used each language on a daily basis.

Given that a bilingual advantage has been demonstrated in inhibition and other aspects of executive control regarding the sorting of information (Bialystok, 1999), we predicted that bilinguals who reported using their languages on a regular basis would outperform monolinguals in span length in the Corsi task, with the greatest advantage emerging in the backward span version. We further predicted that this bilingual group would respond more efficiently (faster, more accurately, or both) than monolingual peers in the Simon condition that included conflicting stimuli and in the condition with more rules.

CHAPTER II

METHOD

Participants

In the current study, we attempted to recruit 200 young adult participants. A power analysis using G*Power (Faul et al., 2007) indicated that a total sample size of 200 would be sufficient for a small effect size of .10 and power of .85. To perform the analysis a priori, we chose ANOVA: Repeated measures, within-between interaction. Alpha was set to .05, and number of groups for all analyses was 2 (bilingual, monolingual). For the Corsi task, we calculated the necessary sample size for 2 measurements (backward span, forward span). For the Simon task, the same parameters were kept, but the number of measurements was set to 6 (neutral-no rule sorting, neutral-rule sorting, incompatible-rule sorting, incompatible-no rule sorting, compatible-rule sorting, compatible-no rule sorting). These parameters indicated that a total sample size of 110 would be sufficient for the Simon task. Because all participants would be performing both tasks, the higher number of 200 was selected as the target sample size to fulfill the needs of both tasks. Although some of the participants in the current study know more than two languages, for clarity the entire sample of participants knowing more than one language will be referred to as "bilingual" regardless of how many languages are known.

We chose ANOVA to more closely align with past research with which we are aiming to compare results. In similar studies, sample sizes of 40-100 participants is common. A sample size of 100 monolinguals and 100 bilinguals with regular use of their languages would meet the criterion of a total sample of 200 for the Corsi task, while going above and beyond the numbers that past researchers have examined. Given the timeline of the current study, and the nature of the bilingual group, we were unable to obtain a sample of 100 bilingual participants. Instead, we obtained a final sample of 158 total participants (111 monolinguals, 47 bilinguals). The sample of 47 bilinguals included 22 participants who know two languages, and 25 participants who know three languages. Participants in each group were between 19-35 years old (*M*=28.68, *SD*=4.3, 7 participants declined to give their exact age.); 93 participants were male, 62 were female, and 3 declined to specify sex.

Participants were recruited through Amazon Mechanical Turk (MTurk) (Amazon Mechanical Turk, Inc., 2018). MTurk was chosen as the platform on which to perform the experiment because MTurk allowed us to obtain the required number of participants in a specific demographic that would otherwise be difficult to access. Researchers have argued that samples from MTurk can be more representative of the general population (Berinsky et al., 2012) and attend to tasks much better than lab samples (Hauser & Schwarz, 2016).[1] Participants from both groups were excluded if their data showed

[1]Researchers could directly explore the quality of data collected through MTurk by comparing with a sample of participants who perform the same tasks in laboratory; however, a comparable in-person sample was outside the scope of the current thesis research study.

patterns of insincerity, such as more than 10 responses less than 100 ms, 5 or more timed out/incorrect responses in a row, 5 or more timed out responses in one condition, or 20 or more timed out/incorrect responses total. Such patterns suggest that a participant is pressing buttons as quickly as possible, or are not fully engaged in the task. Participants were also excluded if they reported having impairments such as color blindness, speech delay, or hearing loss, as this could have an effect on responses to the tasks. These were a priori exclusion criteria.

Bilingual participants were recruited through MTurk using a two-phase procedure. First, a questionnaire that included a Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007) was opened to participants who reported speaking English and who reported speaking two or more languages. It was our goal to reach 1,000 participants with the LEAP-Q because we estimated that having 10 times the 100 necessary bilingual participants take the LEAP-Q would ensure that we received responses from a sufficient number of bilinguals to achieve 100 bilinguals who used their languages on a regular basis. The total number of participants who appropriately responded to the LEAP-Q was 686. Of these 686 LEAP-Q participants, 96 bilinguals who used their languages on a regular basis were identified (i.e., those who did not use any one language more than 60% of the time). MTurk participants were compensated $.35 for completing this phase, which took an average of 4 minutes to complete. This phase of the experiment was named “Language Experience Questionnaire” to ensure that participants did not skew their answers to the survey to fit the needs of the study. Participants who have regular use of their languages were defined as individuals who speak two or more languages and use one language no more than 60% of the time. During the second phase of online recruitment, participants who were regular use bilinguals were invited to participate in the cognitive tasks. MTurk participants received compensation of $2 for completing the cognitive tasks, which took participants an average of 10 minutes. The second phase of the experiment included demographic questions and handedness questions as well as the Simon and Corsi tasks; 96 participants were invited to this phase of the experiment, 40 declined the invitation or did not respond, 9 were excluded due to insincere data or impairments, and 47 responded appropriately to the tasks.

Monolingual participants were also recruited through MTurk; however, it was not necessary for the monolingual participants to complete the LEAP-Q. Monolingual participants only participated in the second phase of the experiment, which included the cognitive tasks, as well as the demographic and handedness questions. Only MTurk participants who indicated that they only speak one language and that they speak English were included. 131 total monolinguals responded to the tasks; however, 20 participants were excluded for insincere data or existing impairments.

Procedure

The cognitive tasks were conducted using the PsyToolkit platform (Stoet, 2010; 2017), and were modified versions of the Corsi and Simon tasks found in the PsyToolkit library. The tasks and each condition were counterbalanced so that half of the participants were presented with the Corsi task first, and half of the participants were presented with the Simon task first. The Corsi task (Berch et al., 1998) had two conditions, a forward recall condition and a backward recall condition. Each participant was presented with an array of nine purple squares on the computer monitor. One of the squares appeared to "light up" by turning yellow and remained yellow for 300 ms. After a 300 ms delay, another square turned yellow and remained for 300 ms. The participant was prompted to click the boxes in the correct order in which the yellow squares appeared. To discourage the participant from trying to click the squares too soon, the mouse cursor was turned off at the start of the trial until 100 ms after the prompt to begin. After checking the boxes, the participant clicked a green square at the bottom right corner of the screen labeled "done" to indicate that they had finished that trial. If the participant checked the correct boxes, a yellow smiley face emoji appeared over the "done" button for 500 ms. If the participant recalled the boxes incorrectly, a red sad face emoji appeared over the "done" button for 500 ms. Each trial had a new randomly generated array of nine purple squares. Participants were shown a span starting with two squares that lit up yellow, with two possible trials for each span length. If the participant correctly reproduced the span for one of the trials, then the participant progressed to the next span with a new array of purple squares, and the span to be recalled increased by one more square lighting up. If a participant clicked on an incorrect sequence of squares, they were given a second span of the same length. Once a participant incorrectly recalled both trials for a given span, the task was over. The backward span condition was performed in the same manner, but the participants were instructed to recall the boxes in the opposite order in which they appeared. In this task, the backward span always occurred first, to maximize the difficulty level of the task.

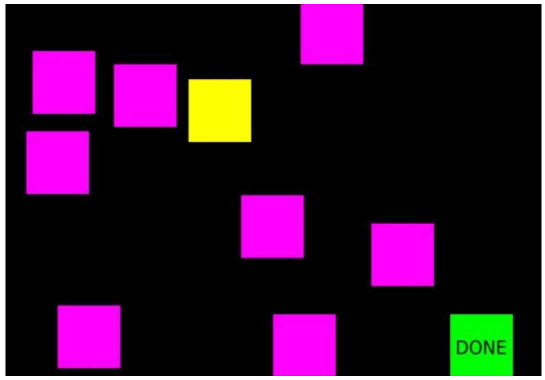

Figure 1. The Corsi task as seen on the participant's computer monitor.

The Simon task (Bialystok, 1999) was comprised of six conditions: Rule sorting (2) x Conflict (3). In the conditions with no rule sorting, participants were presented with two colored square stimuli on a computer screen. Participants received the instructions in a series of dialogue boxes. Each rule for the task was presented in its own dialogue box. For example, "if you see a blue square, press the 'A' key. Press the spacebar to continue." After pressing the spacebar, a new dialogue box was presented to the participant with the next rule, "If you see a yellow square, press the 'L' key. Press the space bar to continue." Each response key had only one stimulus associated with that key. In the conditions with rule sorting, there were four colored squares, with each response key corresponding to two stimuli. The directions for each stimulus were given separately. For example, "If you see a red square, press the 'A' key. Press the spacebar to continue." Followed by "If you see a purple square, press the 'A' key." These instructions were shown in the beginning of each block, and pertained to the entire block. In the rule sorting condition, there were four different stimuli that the participant must learn, with four rules that the participant must "sort" or switch between. This condition was meant to simulate code-switching, during which individuals are able to switch between rules, grammar, and vocabulary, in language. In one trial, a red square indicated that the "A" key should be pressed, but in another trial, a purple square indicated that the "A" key should be pressed. Further, there were two colored squares corresponding to each key, giving a total of four stimuli that must be remembered. This configuration created a task that placed much more demand on EF, as the task now required rule switching and working memory to be engaged in addition to inhibition during the incompatible trials. These two conditions, rule sorting and no rule sorting, were run in separate blocks with the more difficult rule sorting condition always occurring first. By placing the rule sorting condition first (the condition with four stimuli), we ensured that the task placed the most demands on EF.

In the Conflict conditions, there were compatible, neutral, and incompatible trials. In the compatible trials, the stimuli will appear on the same side of the screen the corresponding response key. A red square indicating to push the "A" key, would appear on the left side of the screen, which is congruent with the "A" key being on the left side of the keyboard. In the neutral condition, the stimuli appeared in the center of the screen. This condition acted as a baseline measure and was performed in a separate block from the other two conditions. By asking participants to respond to the stimuli as they appeared in the center of the screen, we were able to assess by RT whether or not the conditions adding two more stimuli or moving the stimuli to either side of the screen, truly placed more demands on EF, as we predicted. By keeping the neutral condition separate from the compatible incompatible trials, we ensured a true baseline with no carry over effect from the more difficult conditions. In the incompatible trials, the stimuli appeared on the opposite side of the screen as their associated response key. A red square indicating to push the "A" key, would appear on the right side of the screen, whereas the

“A” key is located on the left side of the screen. In this way, we created conditions that used both working memory and inhibition to varying degrees. Each square remained on the screen until the participant pressed either the “A” or the “L” key, or for five seconds. If the participant pressed the incorrect key or failed to respond using the “A” or the “L” keys within five seconds, an error message was displayed for 500 ms before moving on to the next stimuli. Once the correct key was pressed, there was a 300 ms delay during which a re-centering cross appeared in the middle of the screen.

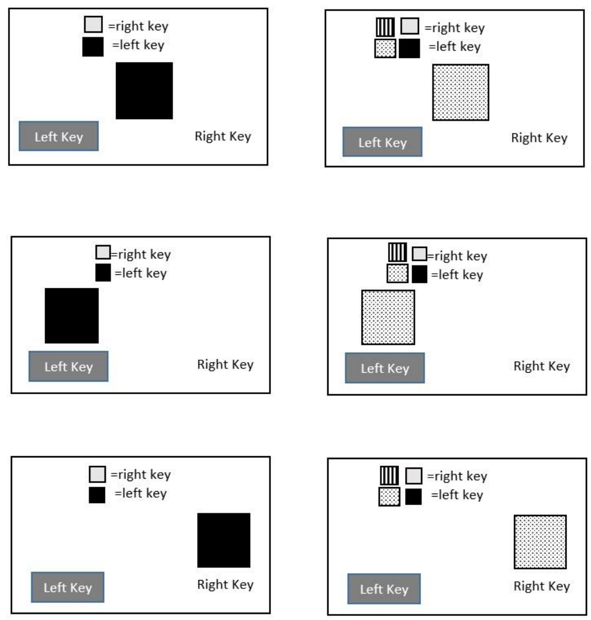

Figure 2. The screen orientation for rule sorting, conflict, and neutral conditions of the Simon task.

The neutral condition was in a separate block, and the compatible and incompatible conditions were tested together. Keeping the neutral condition in its own block had both logistical benefits in programming the tasks, reading the data sent to researchers, and experimental benefits in providing a pure baseline measure that is not tainted by carryover effects of the compatible and incompatible trials. Participants completed one block of thirty neutral trials with four rules, then one block of thirty randomly ordered compatible and incompatible trials with four rules. They then completed another block of thirty neutral trials with two rules followed by the final block of thirty randomly ordered compatible and incompatible trials with two rules.

In order to encourage transparency and participate in open science practices, the proposed study was preregistered on the Open Science Framework (OSF) at https://help.osf.io/hc/en-us (Center for Open Science, 2018).

CHAPTER III

ANALYSES

Dependent Variables

In the Corsi task, we collected data for each trial that a participant executed. Data included the number of items to be recalled, the number of items that were recalled in the previous trial, and whether or not the participant recalled this span. The data corresponding to the number of items to be recalled in the previous span indicated whether this was the first or second trial with the span number. The number of the final span correctly recalled was recorded for each condition and used for data analysis. A final span number was considered to be the highest number of squares that a participant correctly recalled in that condition.

In the Simon task, we collected reaction times (RTs) and accuracy across all six conditions for each participant. RTs were measured in milliseconds and defined as the time between the onset of a stimulus and the onset of a participant's key press. In our final analyses, we used participants' mean scores in each condition. Accuracy was the number of correct first key pushes each participant made within a condition. If a participant pushed an incorrect key followed by a correct key, the trial was scored as incorrect. Because the number of compatible and incompatible trials varied by participant, to ensure true trial randomization, percentages were calculated by dividing the number of correct keystrokes by the number of trials the participant received in that condition. These percentages were used in final analyses.

Analyses

For the Corsi task, a 2 (Language Status: monolingual, bilingual) x 2 (Span Direction: forward, backward) ANOVA was performed to examine the main effects of Language Status and Span Direction and the interaction between Language Status and Span Direction. Forty monolinguals and 11 bilinguals were excluded because they had a score of zero in one or both of the spans. These exclusion rules were decided a priori. The high number of exclusions could stem from participants' confusion about when to respond with a backward sequence and when to respond with a forward sequence. A main effect of Language Status was predicted, such that bilinguals were expected to have larger recall spans than monolinguals. Monolinguals had a mean span of 5.22, and bilinguals had a mean span of 5.37. There was no main effect of language, $F(1, 103) = .39$, $p = .53$, $\eta_p^2 = .004$. A main effect of Span Direction was predicted; such that larger recall spans were expected in the forward than the backward condition. The forward span had a mean of 5.65, and the backward span had a mean of 4.9. As expected, a main effect of Span Direction was found, $F(1, 103) = 20.25$; $p < .001$, $\eta_p^2 = .164$. Finally, an interaction between Language Status and Span Direction was predicted, such that the difference between the bilinguals and the monolinguals was expected to be greater in the backward span condition than in the forward span condition. In the forward span, there was a mean difference of .12 between monolinguals and bilinguals. In the backward span, there was a larger mean difference of .54, with bilinguals having a greater span. The interaction between Span Direction and Language Status approached significance, $F(1, 102)= 3.82$; $p= .053$, $\eta_p^2= .036$. A planned comparison *t*-test also showed that the difference between the bilinguals and monolinguals in the backward span was approaching significance $t(103)= -1.6$, $p= .056$, $d=.33$.

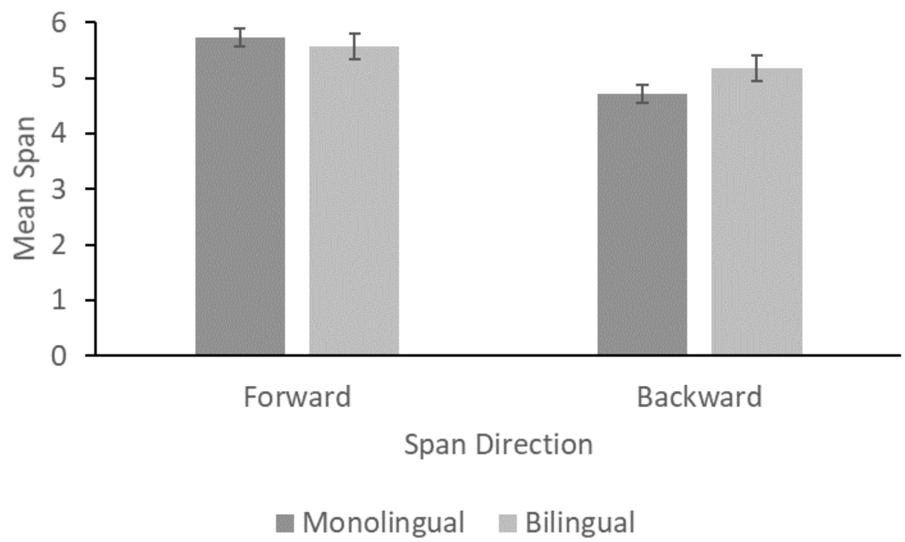

Figure 3. Mean span recall for monolinguals and bilinguals in backward and forward Corsi span. Error bars represent standard error.

For the Simon task, two separate 2 (Language Status: monolingual, bilingual) x 2 (Rule Sorting: 2 Stimuli, 4 Stimuli) x 3 (Conflict: Compatible, Neutral, Incompatible) mixed ANOVAs were performed, one for accuracy and one for RT. The mixed ANOVAs allowed us to examine the main effects and interactions of Language Status, Rule Sorting, and Conflict on accuracy and RT.

Outliers for RT, defined as participants who received scores greater than three standard deviations from the grand mean (M= 641.92, SD= 267.82), were excluded from the analyses of RT. These exclusion criteria were a priori and preregistered. A total of 10 (8 monolingual, 2 bilingual) participants were excluded as outliers from RT analyses. A main effect of Conflict was predicted, such that the incompatible position would have the slowest RT. Conditions that placed the stimulus in a compatible position had a mean RT of 588.6 ms. Conditions that placed the stimulus in a neutral position had a mean RT of 586.8 ms. Conditions that placed the stimulus in an incompatible position had a mean RT of 624.5 ms. There was a mean difference of 37.74 ms between incompatible and neutral ($p < .001$) and a mean difference of 35.94 ms between incompatible and compatible ($p <$ .001.) There was no significant difference between the neutral and compatible stimulus positions. A main effect of Conflict was found for RT, $F(2, 145) = 18.38$, $p < .001$, $\eta_p^2 = .112$. As predicted, trials that placed the stimulus in an incompatible position with the response key were significantly slower than compatible and neutral trials. A main effect of Rule Sorting was expected such that conditions with four rules would have slower RT than conditions with two rules. Conditions with four rules had a mean of 646.2 ms, and conditions with two rules had a mean of 553.8 ms. There was a mean difference of 92.41 ms ($p < .001$) between trials that had four rules to remember and trials that had only two rules, with four rule trials being significantly slower. As expected, we found a main effect of Rule Sorting, $F(1, 146) = 164.22$, $p < .001$, $\eta_p^2 = .53$. The mean for bilinguals was 594 ms and the mean for monolinguals was 606 ms. There was no main effect of Language $F(1, 146) = .122$, $p = .73$, $\eta_p^2 = .001$. There was a significant interaction between Language and Rule Sorting, $F(1, 146) = 6.82$, $p = .01$, $\eta_p^2 = .045$, such that bilinguals had RTs that were faster than monolinguals in conditions that had only two rules. There was a significant interaction between Rule sorting and Conflict, $F(2, 145) = 3.12$, $p = .046$, $\eta_p^2 = .021$, such that trials that placed the stimuli in an incompatible position had slower RTs from neutral position when there were four rules, and when there were two rules it was compatible trials that had the fastest reaction times. Planned comparisons *t*-tests showed that while the differences are not statistically significant between monolinguals and bilinguals across Conflict conditions, there is a small effect size indicating that a larger bilingual sample size more equal to the monolingual sample could show a significant difference. For compatible trials, the mean difference in RT between monolinguals and bilinguals was 8.3, $t(120.5) = .28$, $p = .4$, $d = .04$. For the neutral conditions, the mean difference in RT between monolinguals and bilinguals was 16.3, $t(123.6) = .545$, $p = .3$, $d = .083$. For incompatible trials, the mean difference in RT between monolinguals and bilinguals was 11.0, $t(110.8) = .345$, $p = .4$, $d = .06$. For trials with two rules, the mean difference in RT between monolinguals and bilinguals was 30.7, $t(116.6) = 1.0$, $p = .16$, $d = .16$. For trials with Four rules, the mean difference in RT between monolinguals and bilinguals was 6, $t(116.9) = .23$, $p = .42$, $d = .04$.

Table 1
Mean RTs across conditions for monolinguals

	Number of Rules		
Conflict Condition	Two	Four	Difference
Compatible	558	628	70
Neutral	565	625	60
Incompatible	585	675	90

Table 2
Mean RTs across conditions for bilinguals

	Number of Rules		
Conflict Condition	Two	Four	Difference
Compatible	522	646	124
Neutral	534	624	90
Incompatible	559	679	120

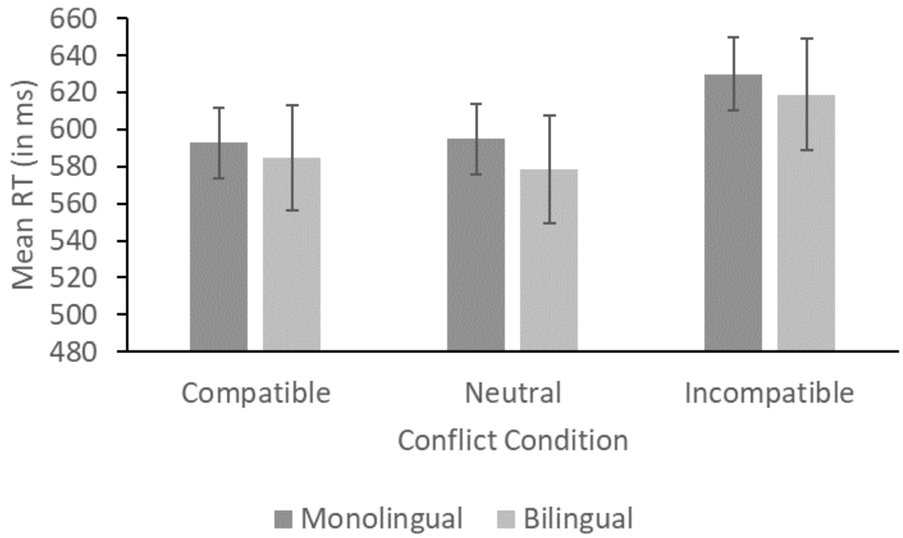

Figure 4. Mean reaction times for monolinguals and bilinguals across Conflict conditions. Error bars represent standard error.

Figure 5. Mean reaction times across Rule Sorting conditions. Error bars represent standard error.

Nine monolingual outliers were excluded from the accuracy analyses for exceeding three standard deviations from the grand mean (M= .94, SD=.1). These exclusion criteria were decided a priori. As predicted, the trials in which the position of the stimulus was compatible with the response key position had the greatest accuracy, with a mean of 96.8%. The neutral trials had a mean accuracy of 94. 9% and the incompatible trials had the lowest accuracy with a mean of 92%. A main effect of Conflict was found $F(2, 146)= 36.18$, $p< .001$, $\eta_p^2= .198$. These data suggest that although compatibility did not facilitate RT, accuracy was facilitated. For accuracy, trials that had four rules had a mean of 93.9% correct, and trials that had two rules had a mean of 95.2% correct. A main effect of Rule Sorting was obtained, $F(1, 147)= 8.13$, $p= .005$, $\eta_p^2= .05$. Trials that had only two rules were 1.3% more accurate than trials that had four rules (p=.005.) Bilinguals had an overall mean of 95.1% correct and monolinguals had a mean of 94.1% There was no main effect for Language in accuracy, $F(1, 147)= 1.9$, $p= .171$, $\eta_p^2= .013$. Planned comparisons *t*-tests showed that while the differences are not statistically significant between monolinguals and bilinguals across Conflict conditions, there are small effect sizes indicating that a larger bilingual sample size more equal to the monolingual sample could show a significant difference in some conditions. For compatible trials, the mean difference between monolinguals and bilinguals in accuracy was .003, $t(147)= .366$, $p= .36$, $d= .06$. For the neutral conditions, the mean difference in accuracy between monolinguals and bilinguals was .005, $t(147)= .66$, $p= .26$, $d= .12$. For incompatible trials, the mean difference in accuracy between monolinguals and bilinguals was .02, $t(147)= .345$, $p= .05$, $d= .29$. For trials with two rules, the mean difference in accuracy between monolinguals and bilinguals was .01, $t(147)= 1.5$, $p= .07$, $d= .26$. For trials with Four rules, the mean difference in accuracy between monolinguals and bilinguals was .009, $t(147)= .96$, $p= .17$, $d= .17$.

Table 3
Mean accuracy scores across conditions for monolinguals

	Number of Rules		
Conflict Condition	Two	Four	Difference
Compatible	96.7%	96.6%	0.1
Neutral	95.8%	93.5%	2.3
Incompatible	91.4%	90.4%	1

Table 4
Mean accuracy scores across conditions for bilinguals

	Number of Rules		
Conflict Condition	Two	Four	Difference
Compatible	97.3%	96.6%	0.7
Neutral	96.1%	94.3%	1.8
Incompatible	93.8%	92.4%	1.4

Figure 6. Accuracy across rule sorting conditions. Error bars represent standard errors.

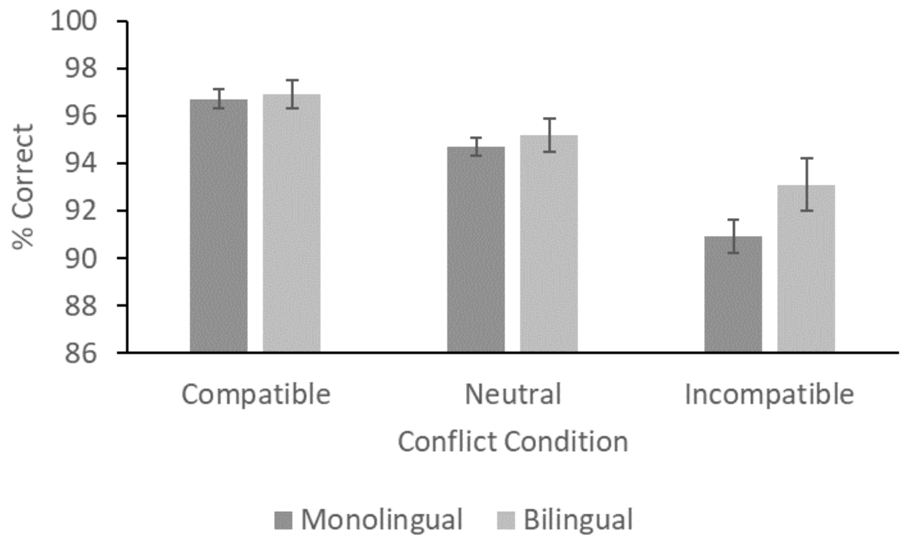

Figure7. Mean accuracy for monolinguals and bilinguals across conflict conditions. Error bars represent standard errors.

Although there was no significant main effect of Language for accuracy or RT, and comparisons between conditions should remain tentative until full power for the bilingual group is reached, the predicted patterns are emerging. Bilinguals were consistently more efficient in all three Conflict conditions, with this efficiency being most apparent in the accuracy of incompatible trials. The expected patterns are also exhibited in the data from the Corsi task, with the bilingual group performing more efficiently than the monolingual group in the backwards span.

CHAPTER IV

DISCUSSION

Overall Bilingual Advantage

It is important to remember that all discussions and conclusions regarding comparisons between bilinguals are tentative, given that the bilingual group is currently underpowered. In the Corsi task, we expected that bilinguals would remember longer forward spans than monolinguals. We further expected that this effect would be greater for the backward span, because a backward span places more demands on WM. A larger forward span recall for bilinguals would support the hypothesis that an overall bilingual advantage in EF exists. The greater effect in the backward span would support the hypothesis that equal use bilingualism gives bilinguals the experience necessary to learn how to navigate greater demands on WM. While differences between bilinguals and monolinguals were not evident in the forward task, a pattern emerged showing that bilinguals remembered higher spans in the backward task. This pattern suggests that bilinguals may be able to better navigate greater demands on WM.

The Simon task was expected to show an overall bilingual advantage across domains of EF. Because all domains of EF are hypothesized to work together (Miyake et al., 2000), we expected that by placing a demand on one function, such as switching, working memory, or inhibition, every domain would be affected. We further expected that greater demands across domains of EF would require greater processing times to negotiate such demands. An overall faster RT in the bilingual group would have suggested a greater capacity for negotiating such demands. The greatest overall advantage in the Simon task was expected to be observed in the condition placing the greatest demands on EF, the condition with four rules and incompatible stimuli placement. If the demands placed on each EF domain affect all domains, then placing demands on multiple domains would create a situation of high EF demands in which faster, more efficient negotiations of such demands would be the most beneficial.

While there was no main effect of Language, the expected patterns within the data were evident, and it is possible that these effects will become more pronounced with a sufficiently powered bilingual sample. Bilinguals showed a pattern of responding faster than monolinguals across all three Conflict conditions, although it is important to reiterate that these differences were not statistically significant with this sample size. Bilinguals were significantly faster than monolinguals in trials that had two rules, whereas trials that had four rules were not statistically different. This interaction will be further explored with a larger bilingual sample, because we predicted that bilinguals would have faster RTs in both rule sorting conditions, and that the difference would be greater in the four rules condition. Having a larger bilingual sample size should allow us to see a clearer picture of this interaction.

Similar patterns could be observed for accuracy in the Simon task. Bilinguals were trending toward greater accuracy than monolinguals across both rule sorting conditions. Bilinguals were also trending toward greater accuracy across all three

Conflict conditions, with the biggest difference in the incompatible condition. A significant difference in accuracy would have indicated that a bilingual advantage is most prominent in contexts that place a lot of demand on EF. When the target bilingual sample size is obtained, it could show that the differences between bilingual and monolingual groups are significant in high demand conditions.

While exploring only one level of language use may be considered a drawback of the current study, doing so allows a link to be established between regular language use and an advantage in EF during young adulthood. With evidence of an advantage in this population, future studies could explore language use as a continuous variable rather than focusing only on bilinguals who regularly use their languages. Such research would further support the idea that language experience has varying effects on the bilingual advantage in EF and provide an additional explanation for variations in results across studies. If a bilingual advantage in EF is obtained in samples with homogenous language experiences, then perhaps the advantage is being diluted (or eliminated) in samples in which language experiences or context are considered by including some participants with – and some participants without – a bilingual advantage.

Advantage in Navigating Demands

An overall bilingual advantage as discussed above would provide insight into how bilingual experience affects cognitive abilities in EF. Further insight could be gained from comparing group performance as participants move from tasks with fewer demands on EF to tasks with increasing demands on EF. In the Corsi task, participants were expected to have overall lower scores in the high demand backward span than in the low demand forward span. However, a smaller difference between the backward and forward span scores for bilinguals would indicate better adaptation to the high EF task. A larger difference between these scores for monolinguals would indicate that the backward span placed demands on EF that were more difficult for monolinguals to navigate. Indeed, we found a larger difference between the monolingual forward and backward spans than for the bilingual forward and backward spans. These results support the hypothesis that the language experience of regular use bilinguals allows bilinguals to navigate greater demands on EF.

Ensuring Accurate Data

Some of our hypotheses were designed to provide information regarding the validity of the data. We expected that, regardless of evidence for or against a bilingual advantage, scores for the backward Corsi task would be lower than scores for the forward Corsi task. We also expected slower RTs and lower accuracy in the Simon task when there were four stimuli than when there were two stimuli. We further expected slower RTs and lower accuracy in the incompatible condition than in the compatible and neutral conditions. Such results would indicate that the design of the task, and the online experiment platforms, are providing valid data. All of these expectations came to fruition, except for slower RTs in the neutral condition relative to compatible condition; RTs were equivalent in these conditions. This pattern of results indicates that the compatible position of the stimulus to the response key had a facilitating effect on accuracy, but not on RT.

Limitations

Previous studies have yielded conflicting results, causing the bilingual advantage hypothesis in EF to be controversial. While patterns seem to be emerging in the results of the current study that would be consistent with a bilingual advantage, it is important to note that most of the differences are not statistically significant. It is difficult to draw conclusions until the sufficiently powered, originally planned, bilingual sample is obtained. The sufficiently powered sample size could show these patterns are statistically significant; however, it is also possible that a larger sample size will continue to show patterns that are not statistically significant (or that these patterns will disappear completely). Although there could be multiple explanations for null results, perhaps the most parsimonious explanation would be that an advantage in EF for bilinguals simply does not exist. If an advantage does not exist in the bilingual population, then looking for an advantage in specific sub-groups of bilinguals would be unfruitful. Indeed, a recent study that included a large pool of bilinguals broken into subgroups, suggests that no bilingual advantage in EF exists in any group (Dick et al., 2019).

However, there are other explanations that may account for null results, while also being consistent with evidence from previous studies. In the current study, we only considered participants with (approximately) equal language use. Perhaps language use is not the factor in language experience that provides learning opportunities to acquire superior navigation of EF demands. Alternatively, there are many interactions between language experiences, and not just one that can be isolated. Further, if there is no learning process involved, but rather a “priming” of the EF to be ready to perform under demands, then perhaps individual differences do not play a role, and the advantage is solely or jointly dependent on the context of the situation. Perhaps regular use bilinguals have an advantage in inhibition tasks only in dual language contexts in which there are demands on EF. Age may also play a role, as all participants in the proposed study were young adults. The advantage may only be seen during a time of development or decline in EF, such as during childhood or in older adults.

The decision to include participants who speak three languages as well as two languages must also be considered. Doing so allows for a broader participant base; however, the number of languages one speaks could play a role in individual language experience. Juggling three or more languages could be different than juggling just two languages.

The methods in recruiting the monolinguals and bilinguals varied slightly which may have caused some differences in the groups. While the recruiting processes for each group was made as similar as possible, bilinguals were selected and invited to participate in the executive function tasks based on their responses to the LEAP-Q and monolinguals did not receive the LEAP-Q. Both monolinguals and bilinguals were recruited using filter questions to ensure each participant was placed in the correct group. Because of these filter questions and the LEAP-Q, neither group was randomly selected.

The final reason for the null results may be that the tasks used were not sufficiently sensitive to detect an advantage in EF, or that the bilingual group was not sufficiently large to power the study appropriately with these tests. Addressing these concerns in future research may add clarity to the controversy by demonstrating what conditions may result in a bilingual advantage, and why many studies have been unable to obtain supporting evidence. Consideration for language experience and context will be the key, not only for studying cognitive advantages of bilingualism, but also for studying language processing and EF more broadly.

REFERENCES

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics,* 20, 242-275. DOI: 10.1016/j.neuroling.2006.10.003

Alloway, T., Gathercole, S., & Pickering, S. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development,* 77(6), 1698-1716

Amazon Mechanical Turk, Inc. (2018) Amazon Mechanical Turk. Retrieved from: MTurk.com

Berch, D.B., Krikorian, R., & Huha, E.M., (1998) The Corsi Block Tapping Task: Methodological and theoretical considerations. *Brain and Cognition*, 38, 317-338.

Berinsky, A. J., Huber, G. A., & Lenz, G.S., (2012) Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis.* 20, 351-368. DOI: 10.1093/pan/mpr057

Benson, J. E., Sabbagh, M. A., Carlson, S. M., & Zelazo, P. D., (2013) Individual differences in executive functioning predict preschoolers' improvement from theory of mind training. *Developmental Psychology,* 49(9), 1615-1627. DOI: 10.1037/a0031056

Bialystok, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology*, 24(4), 560-567

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development,* 70(3), 636-644

Bialystok, E., Craik, F., Klein, R., & Viswanathan, M., (2004), Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and Aging*, 19(2), 290-303 DOI: 10.1037/0882-7974.19.2.290

Blom, E., Kuntay, A., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, 128, 105-119. DOI: 10.1016/j.jecp.2014.06.007

Center for Open Science. (2018) Open Science Framework. Retrieved from: OSF.IO

Dick, A. S., Garcia, N. L., Pruden, S. M., Thompson, W. K., Hawes, S. W., Sutherland, M, T., Riedel, M. C., Laird, A. R., & Gonzalez, R. (2019). No evidence for a bilingual executive function advantage in the nationally representative ABCD study. *Nature Human Behavior*, 3, 692-701. DOI:10.1038/s41562-019-0609-3

Faul, F., Erdfelder, E., Lang, AG., & Buchner, A. (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39 (2), 175-191. DOI:10.3758/BF03193146

Gordon, K.R., (2016). High proficiency across two languages is related to better mental state reasoning for bilingual children. *Journal of Child Language*, 43, 407-422. DOI: 10.1017/S0305000915000276

Green, D.W, (1998) Mental control of the bilingual lexico-semantic system. *Bilingualsim: Language and cognition.* 1, 67-81. DOI:10.1017/S1366728998000133

Hauser, D. J., & Schwarz, N. (2016) Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods.* 48, 400-407. DOI: 10.3758/s13428-015-015-0578-z

Incera, S., & McLennan, C. T. (2018). Bilingualism and age are continuous variables that influence executive function. *Aging, Neuropsychology, and Cognition*, 25(3), 443-463

Luk, G., DeSa,E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, 14(4), 488-495

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M., (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals, *Journal of Speech, Language, and Hearing Research,* 50(4), 940-967

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. Cognitive Psychology, 41(1), 49– 100. DOI:10.1006/cogp.1999.0734

Morales, J., Alejandra, C., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology,* 114(2), 187-202

Paap, K. R., & Greenberg, Z. I., (2013), There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology,* 66(2), 232-258.

Paap, K.R., Johnson, H. A., &Sawi, O., (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex,* 69, 265-278

Ratiu, I., & Azuma, T. (2015). Working Memory Capacity: is there a bilingual advantage? J*ournal of Cognitive Psychology*, 27, 1–11. DOI: 10.1080/20445911.2014.976226

Rubio-Fernandez, P. (2017). Why are bilinguals better than monolinguals at false-belief tasks? *Psychonomic Bulletin & Review*, 24(3), 987-998. DOI: 10.3758/s13423-016-1143-1

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42(4)*, 1096-1104

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31

Warmington, M., Kandru-Pothineni, S., & Hitch, G (2018). Novel-word learning, executive control and working memory: a bilingual advantage. *Cambridge University Press.* DOI:10.1017/S13667289180004

APPENDIX

Notice of Funding

This research was funded by a Graduate Student Research Award from the Office of Research at Cleveland State University.