The Effects of Multilingualism of Executive Function

Isabella Catalano

University of Nebraska - Lincoln

Follow this and additional works at: https://digitalcommons.unl.edu/envstudtheses

Part of the Environmental Education Commons, Natural Resources and Conservation Commons, and the Sustainability Commons

https://digitalcommons.unl.edu/envstudtheses/241

This Article is brought to you for free and open access by the Environmental Studies Program at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Environmental Studies Undergraduate Student Theses by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
THE EFFECT OF MULTILINGUALISM ON EXECUTIVE FUNCTION

by

Isabella Catalano

AN UNDERGRADUATE THESIS

Presented to the Faculty of
The Environmental Studies Program at the University of Nebraska-Lincoln
In Partial Fulfillment of Requirements
For the Degree of Bachelor of Arts

Major: Environmental Studies and Psychology

Under the Supervision of Dr. Anne Schutte and Dr. Marc Goodrich

Lincoln, Nebraska
December, 2018
ABSTRACT

In today's world, monolingualism is in the minority (Alonso et al., 2017); however, there is still a lack of understanding about the potential effects of being bi- or multilingual, and whether there is an effect of bilingualism in executive function is debated, given multiple contradictory studies (Paap et al., 2015). This study aims to more closely examine whether the number of languages spoken is related to executive function. In this study, sixty-three participants (mean age = 19.9 years, males = 10) completed the Stroop and flanker tasks, measures of inhibitory control, as well as the Language Experience and Proficiency Questionnaire, and the Expressive One-Word Picture Vocabulary Test (EOWPVT). There were 23 monolingual participants, 30 bilinguals, and 10 multilinguals. ANOVAs were used to identify potential differences in inhibitory control across monolingual, bilingual, and multilingual participants. Overall, no effects or interactions were found, which is in line with some of the literature but contradicts other studies.
Worldwide, monolinguals are in the minority; the majority of people are in fact bi- or multilingual (Alonso et al., 2017). Multiple languages are always active at once, as evidenced by types of cross-language interactions (Kroll et al., 2012). These interactions include code switching, in which speakers of multiple languages switch between languages during use, and the bidirectional effects of second language (L2) acquisition, in which not only does the first language (L1) affect the L2, but also vice versa (Kroll et al., 2012). This occurs also amongst individuals learning a third language (L3), though there is evidence that L3 acquisition is distinct from L2 acquisition (Alonso et al., 2017, de Bot & Jaensch, 2015), and persists both across similar and dissimilar languages (Kroll et al., 2012). Code switching, in particular, is not a random process: bilinguals do not, for example, code switch when people they are speaking to are not able to understand both languages (Kroll et al., 2012). This implies that some level of cognitive control is needed at all times to prevent spontaneous switching; the degree to which this level of control is generalizeable to other aspects of a bi- or multilingual's life is a matter of much debate at present.

It is fairly well documented that multiple languages influence the brain on a biological level. Bilingualism appears to alter grey matter structure in the inferior frontal and parietal regions while altering white matter connectivity in the anterior corpus callosum (Stein et al., 2014). Bilingualism also encourages activity in the left putamen, particularly in less-fluent languages (Abutalebi et al., 2013), and leads to stronger connectivity between the right and left inferior frontal gyrus in bilinguals (Berken et al., 2016). However, there is also some evidence that language also alters functioning on the psychological level—that is, maintaining multiple languages changes the way people think.
A few negative examples of bilingualism's effects have been found: individuals who are bilingual regularly showed lower overall vocabulary and increased lexical retrieval times compared to monolinguals (Bialystok & Craik, 2010). On the other hand, there have also been examples of positive psychological changes that can occur with bilingualism. For example, bilinguals/multilinguals tended to start displaying symptoms of dementia later in life than monolinguals, and were diagnosed later (Craik et al., 2010). Bilinguals also performed better on memory tasks that were based on executive control, though they did not show any advantages in working memory alone (Baumgart & Billick, 2018). Additionally, bilingual children demonstrated better metalinguistic awareness, the sense of grammar and syntax, than monolingual children when the task involved executive control, such as ignoring a false statement when determining whether or not a sentence was grammatically correct (Baumgart & Billick, 2018). Executive function overall, in fact, has been found to be related to many positive bilingual advantages.

Executive function includes abilities such as working memory, cognitive flexibility, and inhibitory control (Diamond, 2014). In bilinguals, the effect of language ability on inhibitory control has been particularly well-studied. According to Green's (1998) inhibitory control model, bilinguals use inhibitory control to prevent themselves from speaking more than one language at once, since, as previously discussed, all languages are always active. This constant practice strengthens the inhibitory control mechanism, which can then be applied to other situations that require behavioral self-control. Research on bilingual inhibition has used a variety of tasks, and, in the process generated contradictory findings.

Tasks and findings. In the Tower of Hanoi task, bilinguals who code switched less performed better (Festman et al., 2010). This was thought to be because the increased inhibitory
effort required to maintain a single language rather than code switching would translate to a stronger inhibitory response in the Tower of Hanoi task, where incorrect or less efficient options would need to be inhibited (Festman et al., 2010).

Another such task is the Simon task, which uses congruent and incongruent stimuli to test the ability to inhibit response. Bialystok et al. (2008) found an interaction between age—younger and older adults—and language ability—monolingual or bilingual—with older bilinguals demonstrating a more robust inhibitory response than similarly-aged monolinguals. Another study in older adults by Goral et al. (2015) found that balanced bilinguals showed an increase of the Simon effect, the difference in response times between congruent and incongruent trials, with age, while unbalanced bilinguals did not; in other words, older balanced bilinguals performed worse. Though this might seem counterintuitive, the authors proposed that when an individual had to more frequently use their inhibitory control mechanisms, such as when lower proficiency in one language required more effort to inhibit the more dominant language, the inhibitory mechanisms were stronger and therefore protected against a decline in function and resulting increase in Simon effect (Goral et al, 2015). However, there have been some studies that have found no such bilingual advantage. In de Bruin et al. (2015), no effect of bilingualism was found on Simon task reaction times (RT).

The attentional network task (ANT), which tests various aspects of attention including executive control, was used by Costa et al. (2008) to test for the bilingual advantage, and they found that bilinguals were faster and better at resolving conflicting situations than monolinguals.

The Stroop task (Stroop, 1935) measures the inhibitory aspect of executive function. The Stroop task involves comparing reading speeds of two different lists of words. One list is color words printed in ink the same color as the word itself—for example, "red" would be printed in
red ink. A second list uses the same color words, but the ink color is incongruent with the printed word—for example, "red" would be printed in blue ink. In both cases, participants are instructed to only say the color of the ink. This becomes significantly harder in the incongruent setting, because the automaticity of reading interferes with the task of naming ink color. The Stroop effect is the difference between reaction times on congruent and incongruent trials. In several studies, bilinguals performed better than monolinguals on the Stroop task, meaning their reaction time was not as slowed down by the incongruent list as for monolinguals. Heidlmayr and colleagues (2014) found a bilingual advantage that was modulated by immersion environment—when bilinguals spent more time in their L2 environment, and thus had to exert the inhibitory effort of suppressing their L1, they displayed a greater advantage. Bialystok et al. (2008) found main effects of language group, bilingual or monolingual, on the Stroop effect, with bilinguals demonstrating an advantage. One potential confound is that participants whose dominant language is not the same language as the testing language might experience skewed results due to decreased automaticity of reading; according to Rosselli et al. (2002), Spanish-English bilinguals, when balanced (according to scores on the Boston Naming Task), showed little effect of language of administration, but unbalanced bilinguals showed a smaller Stroop effect in their dominant language. Kang and Lust (2018), likewise, showed that bilingual proficiency but not, in contrast to some previous research, code switching, impacted results—regardless of how much an individual code switched, greater proficiency was correlated with better performance on the Stroop task. However, as with the Simon task, not all results are consistent. Kousaie and Phillips (2012) did not find an effect of language group on Stroop task performance.

This pattern of results persists in yet another task, the flanker task (a number of variants exist, originating from Eriksen & Eriksen, 1974), which is similar to the Stroop task in that it is a
measure of inhibitory control that uses of reaction time. However, instead of word color, in the flanker task participants must indicate the direction of a target arrow. The target arrow is either flanked by arrows that are facing the same direction or a different direction—thus, the flanking arrows have to be inhibited. While the concept is similar to the Stroop task, the flanker task is nonverbal, eliminating the language-dominance problem. In several studies, bilinguals have shown an advantage over monolinguals. In Chung-Fat-Yim et al. (2018), bilinguals performed better than monolinguals on the flanker task. Thomas-Sunesson et al. (2018) found that more balanced bilingual children performed better on the flanker task than less balanced bilingual children. This is partially supported by Verreyt et al. (2016), who found that balanced bilinguals perform better; but the authors also looked at the role of code switching, and found that only balanced bilinguals who code switched frequently displayed this advantage. Balanced bilinguals who did not switch frequently performed similarly to unbalanced bilinguals (Verreyt et al., 2016). Despite this, no significant differences were found between monolinguals and bilinguals in Dong and Xie's (2013) study, and though Jiao et al. (2017) observed a bilingual advantage, it only occurred on a modified flanker that placed a high demand on memory storage, and not on the standard flanker task. Similarly, Costa et al. (2009) found a bilingual advantage in a flanker task that required high levels of monitoring, but not low levels of monitoring, suggesting that bilinguals may only have an advantage when a high level of information processing is required.

This range of findings has generated much debate in the literature, as well as a host of potential reasons to explain the discrepancy. In a study by Branzi et al. (2016), researchers used verbal and non-verbal switching tasks to demonstrate that the inhibitory functions involved in bilingualism might not be the same as those involved in more domain-general uses of executive control. Paap et al. (2015) point out that there is little convergent validity between tasks—that is,
although all of these tasks purportedly measure inhibitory control, actual correlations between performance on the different tasks are not high enough to generate confidence that the same construct is being measured in all of the tasks. Kalamala et al. (2017) go a step further: based on event-related potential measurements, their study suggests that selective attention rather than response inhibition might be driving flanker task results. Paap et al. (2015) also suggest that, because the sample sizes of many of these studies number in the twenties, significant effects may be inflated.

*Critiques of the literature.* Other critiques focus more on the role that participants play in the study. Morton and Harper (2007), claim that differences in socioeconomic status (SES) maybe the drivers of bilingual advantage, and showed such in their study. Paap and Greenburg (2013) point out that in many prior studies, bilinguals came from higher socioeconomic backgrounds, and thus may not have been generalizeable to a wider bilingual population. More recently, however, a study which included only participants from a low SES still found a bilingual advantage (Thomas-Sunesson et al., 2018). Culture and immigration may also influence the restuls. Immigration is sometimes confounded with bilingualism (Paap et al., 2015), but also results in a distinct pattern, wherein first-generation immigrants are less likely to be multilingual than second-generation, but third-generation immigrants are likely to be monolingual in the language of the new setting than their parents (Baumgart & Billick, 2018). Cultural differences likely play into this as well. In a study of Chinese-American children, Chen et al. (2014) found that higher Chinese literacy was associated with more accurate performance on an inhibition response task. The authors suggest that, as children become more acculturated (that is, have higher English proficiency), they also adopt a faster but less correct US American approach to these kinds of tasks (Chen et al., 2014).
Similarly to the effects of culture, age of acquisition (AoA) has also been found to impact the role of bilingualism in executive function. Yow and Li (2015) found that AoA for an L2, where younger AoA is associated with more balanced bilingualism, could predict performance in some executive control tasks, notably the Stroop test. Luk et al. (2011) found that AoA affected performance on the flanker task, with early bilinguals performing better than late bilinguals and monolinguals, who performed comparably. Although relative levels of proficiency in an L1 or L2 are important to consider when evaluating a potential bilingual advantage in executive function, there is no consensus on how to define or measure bilinguals' language proficiency. Language does not come with a hard cut-off point before which one is not proficient and after which one is fluent; additionally, due to the vast range of languages that participants may speak, determining proficiency with a laboratory-administered test proves difficult (Luk & Bialystok, 2013). Thus, many studies rely upon self-report data of a participant's language proficiency (Luo et al., 2010; Paap & Greenburg, 2013). This becomes especially relevant when looking at studies with trilingual or otherwise multilingual participants, as they have a broader range of languages to account for.

Some studies with multilinguals control which languages participants speak, such as Abunawara (1992), whose participants all spoke Arabic, Hebrew, and English. This allowed for a design that tested whether the stimulus language or the response language of the Stroop task influenced the result; stimulus language did not have a significant main effect (Abunawara, 1992). In a study of children, Poarch and van Hell (2012) controlled for English and German, but also looked at trilingual participants who did not all speak the same third language. Overall, they found that bilingual and trilingual children performed comparatively, and both better than monolingual children (Poarch & van Hell, 2012). van Heuven et al. (2011) examined differences
amongst trilingual speakers and readers of either three alphabetic languages (German/English/Dutch), two alphabetic languages and one non-alphabetic (English/Malay/Chinese), or three languages with distinct scripts (Uyghur/Chinese/English), showing that type of script influences between-language Stroop effects. Not all studies were able to control for language, however, which adds another element of uncertainty to the results. Marian et al. (2013) had a diverse range of participant-spoken languages, but tested between- and within-language competition only within trilinguals. Note that three of these four studies include only trilinguals and the other was with children; at the time of writing, I was not aware of any studies that compared multilingual adults to bi- and monolingual adults.

Current study. The intention of the present study is twofold: firstly, given the range of results that occur in this area of research, the purpose of this study was to replicate previous studies that found a bilingual advantage. Secondly, a notable gap exists in the literature, namely that the effects of multilingualism relative to bi- AND monolingualism have not been examined in adult populations. Additionally, many studies on bilinguals include multilinguals but categorize them as bilinguals, or else limit multilinguals to three languages only and exclude the possibility of more. I wanted to see if parsing these details apart would reveal effects of multilingualism that are distinct from bilingualism.

The research questions of this study are as follows:

1. Is there an effect of bi/multilingualism on executive function?
2. Does the nature of the inhibitory control task (i.e., verbal or nonverbal) moderate the effect of bi/multilingualism on executive function?
3. Do multilinguals show increased advantages over bilinguals on inhibitory control tasks?
For question one, I hypothesized that there would be an effect of bi/multilingualism on executive function, with bi/multilinguals performing better than monolinguals; related to question two, I hypothesized that a greater effect would be found with verbal tasks than with nonverbal tasks, as verbal tasks are related to language. And finally, I hypothesized that there would be an increased advantage for multilinguals over bilinguals due to multilinguals needing to inhibit even more languages than bilinguals.

Method

Participants

Sixty-three university students (male $n = 10$, average age $= 19.9$ years) from the University of Nebraska-Lincoln psychology student pool were recruited for the study. They were given course credit as compensation for their participation. Of the sixty-three participants, twenty-three considered themselves monolingual, thirty considered themselves bilingual, and ten considered themselves multilingual. All participants spoke English as this was the language of study administration, but other languages spoken included Spanish ($n = 18$), Mandarin Chinese ($n = 8$), Arabic ($n = 4$), Vietnamese ($n = 3$), Malay ($n = 2$), Karen ($n = 2$), French ($n = 2$), and Cantonese ($n = 2$). Languages spoken by only one participant each included Thai, Korean, Czech, Tajik, Burmese, Japanese, Bahasa Indonesian, Russian, Urdu, Hindi, and Gujarati.

Measures

Executive Function Tasks. Participants were administered two executive function tasks in a counterbalanced manner. One task was the Stroop task, while the other was the arrow flanker task. Both tasks were taken from the Inquisit (Millisecond) test library. In the Stroop task, participants were shown color words (red, blue, green, black) displayed in either the same color
as the word (i.e., the word "red" written in red ink) or in a different color (i.e., the word "red" written in green ink). Participants were instructed to always indicate the color the word was displayed in, rather than the meaning of the word itself. Trials could be either congruent, incongruent, or control. In congruent trials, the color and the word were the same; in incongruent trials, they were different. Control trials were provided in the form of colored rectangles instead of words. There was one testing block of 85 trials per participant, divided across conditions (color and congruency). Reaction time latency, the difference in time between the appearance of the stimulus on screen and the participant’s pressing of a key to indicate response, was used to create the measure of inhibitory control: the difference in average reaction time (RT) between congruent and incongruent conditions of the task provided the Stroop effect.

In the flanker task, participants were shown a display with five arrows in a line. The format was based on that of Ridderinkhof et al. (1997). The central arrow pointed either left or right, and the four "flanking" arrows either pointed the same way (i.e., central arrow pointed right, flanking arrows pointed right) or the opposite way (i.e., central arrow pointed right, flanking arrows pointed left). Participants were instructed to press a key to indicate the direction that the central arrow only was facing. In the flanker task used for this study, participants were tested in four blocks of twenty trials each. They were also given practice blocks of five trials each before the testing blocks, as the responding key varied. Reaction time latency was used as a measure of responding.

On both the Stroop and the flanker task, only correct response trials were used in the analysis.

*Language Experience and Proficiency Questionnaire (LEAP-Q).* A modified version of the LEAP-Q (Marian et al., 2007) was used in this study (see Appendix). The LEAP-Q asked
questions about a participants' exposure to language as well as their self-perceived fluency, such as whether or not they considered themselves monolingual, bilingual, or multilingual. Proficiency was rated on a scale of 1 – 10 in speaking, understanding, and writing; these scores were averaged to come up with a proficiency score. Participants were instructed to include any languages they had studied in school but might not consider themselves fluent in as well, to create a more holistic idea of their language experiences. Marian et al. (2017) determined that the LEAP-Q possessed a high level of internal validity, with "global measures of self-reported proficiency" being "generally predictive of language ability" (p 962).

Expressive One-Word Picture Vocabulary Test. The EOWPVT-4 (Martin, 2011) was used as a measure of expressive vocabulary and is considered a discrete naming task. Two versions were administered: the Spanish-Bilingual edition (Martin & Brownell, 2013) as well as the standard edition. Differences included the order of items probed, as well as the population upon which the test was standardized. The standard edition was standardized on monolingual individuals of various ages, socioeconomic status, and demographics. The Spanish-Bilingual edition was standardized across Spanish and English bilinguals of varying ages and language proficiencies. Despite this, validity measures are similar across versions, and "lend strong support to the validity of the EOWPVT-4 as an instrument for use in evaluating vocabulary abilities across a wide variety of ages" (Martin, 2011, p 61). Participants were asked before the administration of the EOWPVT what their answer to the LEAP-Q question about their self-reported fluency was, and then the appropriate version of the EOWPVT was given based on their answer (i.e., a participant who replied they considered themselves multilingual or bilingual was given the bilingual edition, and participants who replied they considered themselves monolingual were given the standard edition). In both cases, participants were asked to give their answers in
English, regardless of what other languages they spoke, as the researchers could not account for the sheer variety of languages spoken. In this sense, both administrations of the test were monolingual assessments. The use of the bilingual edition was considered more appropriate for non-monolingual participants even if their additional language was not Spanish, as the norms are based on a bilingual sample, and there are differences in vocabulary knowledge across monolinguals and bilinguals, with bilinguals tending to demonstrate smaller vocabulary (Bialystok et al., 2009).

Results

Data was analyzed using IBM SPSS software. Three participants, one bilingual and two multilingual, were excluded from any analyses involving the EOWPVT because their standardized score fell below the lowest possible score (<55). On the Stroop task, one monolingual participant scored <50% trials correct (average across participants was 93%), while one bilingual and one multilingual had Stroop effects over 1500ms (average across participants was 219ms); all three were dropped from Stroop task analyses. One multilingual participant on the flanker task scored 50.5% correct (average across participants was 94%) and was dropped from the flanker task analyses (see Table 1 for descriptive statistics). A Levene’s test of homogeneity of variance was run in order to determine whether the small number of multilingual participants compared to bilingual and monolingual participants was an acceptable number with which to run an ANOVA. Levene’s test for both the Stroop task, $p = .531$, and the flanker task, $p = .197$, showed that variance was statistically equal, and so multilinguals were analyzed in this study despite a smaller number of participants.
Monolinguals vs. bi/multilinguals. Participants self-reported whether they considered themselves monolingual, bilingual, or multilingual, and this self-report is what was used to classify them into language groups. For these analyses, multilinguals and bilinguals were considered the same category. Results of the analyses can be seen in Table 2. On the Stroop effect, which is the difference between congruent and incongruent conditions, an ANOVA showed that there was no main effect of language category (monolingual or bi/multilingual) on performance, \( p = .443 \). On Stroop facilitation, the difference between control and congruent conditions, a similar lack of effect was found, \( p = .248 \). Stroop cost, the difference between control and incongruent conditions, was also non-significant, \( p = .822 \).

The flanker effect was measured by taking the difference of the congruent and incongruent trails. The flanker effect for monolinguals versus bi/multilinguals was not significant, \( p = .981 \). Additionally, there were also no interactions between language and standardized score on the EOWPVT (see Table 2) for either the Stroop task or the flanker task. EOWPVT alone was also examined to see if varying levels of English proficiency impacted scores, but again, no significant effects were found.

Verbal vs. nonverbal. To examine the second hypothesis that there would be a difference in the effect of language group across the type of task, verbal (Stroop) or nonverbal (flanker) effect sizes were examined across both tasks. For the Stroop effect, when participants were divided into monolinguals and bi/multilinguals, the partial eta squared was \(.011\); for the flanker effect, the partial eta squared was \(.000\). When participants were divided into three language groups (monolingual, bilingual, and multilingual), partial eta squared for the Stroop effect was \(.012\), while partial eta squared for the flanker effect was \(.008\). While the effect sizes trended
slightly larger for the Stroop than the flanker task, neither was an effect size of any magnitude and so it does not appear as though there is a difference across types of task.

*Monolinguals vs Bilinguals vs Multilinguals.* For this section, participants were divided into three language groups: monolingual, bilingual, and multilingual. No significant main effects of language group were found in the Stroop effect \((p = .727)\), Stroop facilitation \((p = .386)\), Stroop cost \((p = .933)\), or the flanker effect \((p = .807)\). EOWPVT score likewise had no significant main effects, nor was there an interaction between EOWPVT and language group.

Because discrete language groups were not having any effect, data were also analyzed on the basis of proficiency, which was a continuous variable. Proficiency scores were provided on the LEAP-Q, and speaking, writing, and understanding scores were averaged to create a proficiency score for each language. For this analysis, the proficiency scores of a participant's L2 were used; thus, monolinguals had a score of 0. For multilinguals, the most proficient of their additional languages was designated L2. This did not, however, change the results, and there was no main effect for L2 proficiency on the Stroop effect \((p = .261)\), Stroop facilitation \((p = .921)\), Stroop cost \((p = .285)\), or flanker effect \((p = .377)\).

Thus, the lack of effect between monolinguals, bilinguals, and multilinguals still persisted.

**Discussion**

Based on previous research, this study sought to examine further any potential effects that bilingualism and multilingualism would have on executive function, specifically inhibitory control. However, overall no significant effects or interactions were found.
Effect of bi/multilingualism on executive function. No statistical difference was found between monolingual and bi/multilingual performance on either the Stroop task or the flanker task, both of which were measured via the difference in RT between the congruent and incongruent conditions. In theory, bi/multilinguals should have shown a decreased reaction time, as experience with continuously inhibiting one language (Green, 1998) would generalize to other types of inhibition, including needing to inhibit one stimulus while responding to another. However, this was not supported by our data. This is consistent with some previous research that indicates that the bilingual advantage is at best elusive and possibly entirely nonexistent (Costa et al., 2009; Paap & Greenburg, 2013); however, given the amount of research that does show effects of more than one language on executive function, this finding is a little surprising.

Effect of task type. Though the hypothesis going into this study was that a verbal task would display a greater difference in performance than a nonverbal task, because language is involved, this was not supported by the data. Neither task showed any significant effects.

Do multilinguals differ? The third and final research question of this study was 'do multilinguals show increased advantages over bilinguals on inhibitory control tasks?' The hypothesis was that there would in fact be a difference, with multilinguals showing a greater advantage than bilinguals. Even divided into three language groups, the results do not show any significant difference between language groups on either the Stroop task or the flanker task. No effects were found when accounting for differing proficiency levels within the sample, either, as a continuous or a discrete variable. This is contrary to some previous research, such as Heidlmayr et al. (2014), in which it was found that the bilingual advantage was in fact reinforced by additional (trilingual) language use, but also aligns with other research such as that by Poarch and van Hell (2012), who found no difference between bilinguals and trilinguals.
Potential explanations. The number of multilingual participants in the present study was small compared to monolingual and bilingual groups, and this small number should be taken into account when considering the results, despite Levene's test.

Another possible explanation involves the tasks. The flanker task, in particular, appears to have generated much controversy over the bilingual advantage it may or may not show. Paap et al. (2015) note that different versions of the flanker task are not convergent with each other; thus, it is possible that the version used in this study differed from previous studies, and, therefore, did not replicate the results. Struys et al. (2018) was one of many studies to find no bilingual advantage in the flanker task—the authors of that paper, however, suggested that any differences in performance may not be due to language ability, but rather an individual's approach to a task.

This explanation is particularly relevant when combined with culture—the Chen et al. (2014) study found that participants with higher proficiency in Chinese (and thus stronger influence of Chinese culture) showed a different pattern of responding than more English-proficient participants. Namely, those with higher Chinese proficiency tended to respond slower, but more accurately. In a task such as the Stroop or flanker where reaction time is of the essence, this cultural difference could be a potential confound. The population from the present study came from a wide variety of backgrounds, from US American to Chinese to Syrian to Mexican. This heterogeneity of culture, though a boon in that it could allow for greater generalizability, may also mean that cultural difference could cancel each other out or otherwise confound the results.

Another potential explanation, as posited in a review by Hilchey and Klein (2011), is that many of the effects of language groups are found in older participants. Given that the present
study has a young mean participant age at 19.9 years, it is possible that the age range tested did not allow for later-life effects of bi- and multilingualism to be yet apparent.

Finally, domain-general inhibitory control is a process that can be used in many areas of life, not just language. Moreno and Farzan (2015) found that training in music leads to "robust and long-lasting" (p 147) changes in inhibitory control. Another study found that short-term music or language training in children produced similar benefits in executive control (Janus et al., 2016). Participants in this study were not asked to identify whether or not they played an instrument or had musical training, and so this cannot be ruled out as a factor which influenced the results. Additionally, a study by Badzakova-Trajkov et al. (2008) found interference effects of tapping a hand while reading in L1 and L2, with bilinguals showing more left-hand interference than monolinguals. Given that responding on the Stroop task in the present study took place via pressing keys on the keyboard with both hands using the same four keys throughout the testing period, it is possible that handedness interfered with the study (the flanker task did, however, counterbalance responding keys).

Limitations. Executive function involves several cognitive processes, and it is possible that the tests used in this study do not capture every nuance. Weissburger et al. (2015) found that executive control used in language switching was most active in trials that required not switching tasks; in other words, needing to refrain from switching tasks is when executive control is most used. This is an effect not tested by the flanker or Stroop tasks. Thus, it is possible that a major component of what makes up the bilingual advantage was not measured. Also not measured by this study was socioeconomic status; as a result, socioeconomic status could not be controlled for and differing SES across participants might have influenced the results. Likewise, age of acquisition was not examined in this study.
Another drawback of the present study is the lack of multilingual participants, which made certain aspects of the analysis difficult to impossible (i.e., examining performance within the multilingual group to see if the language effect was additive when three and four languages were considered separately). To further complicate the language issue, because of the broad range of languages, participants self-reported proficiency. This generated some confusion amongst participants, some of whom did not know how to classify themselves (be it monolingual, bilingual, or multilingual). This was especially a problem for those participants who spoke another language at home or heard another language being spoken at home, but did not know how to read it, and thus did not consider themselves bi- or multilingual despite rating themselves high on speaking and understanding proficiencies.

Additionally, the language of administration (i.e., the language in which both the executive function tasks as well as the LEAP-Q were given in, which in this case was English) might have been a problem for those bilinguals and multilinguals for whom English was not the most proficient language. Unfortunately, by necessity, the administration of the tests is dependent on the language capacity of the researchers as well—because allowing the tests to be administered in various languages would require a limitation on the languages spoken by participants. While Abunuwara (1992) found no effect of stimulus language on response time, the trilingual sample used in that study was also constrained to three specific languages, a difference from this study, which contained nineteen different languages; this narrowing of languages allowed brings into question a study's generalizability to speakers of other languages and other language combinations.

The same issue holds true for proficiency testing—due to the sheer variety of languages spoken, it was not possible to conduct a proficiency test in every language the participants spoke.
This was a problem on the EOWPVT—when standardized, most scores should be close to or around 100, and a score of <55 should account for less than 1% of the population. Despite this, there were three participants who had scores of <55: this could indicate that the EOWPVT was not accurately measuring this specific population. One possible explanation, related to age of acquisition, is that the EOWPVT is structured so that words typically familiar to children compose the lower levels of the test, while more advanced words form the later portion of the test. However, adult learners of English may learn the more advanced words (particularly if they are learning English by way of attending a university) and skip over the lower level words. This pattern of language learning poses a problem with the way the EOWPVT is administered, in that there is a 6-item ceiling that ends the test even if participants might know more advanced words down the line.

Even within English, regional differences in naming (i.e., US American English "truck" versus British English "lorry") could impact tests of expressive vocabulary such as the EOWPVT, where a limited number of words are considered correct answers. Given that the EOWPVT was normed on US Americans, anyone who learned English in a non-US context faced the additional challenge of American-labeled images.

Conclusion. Overall, the bilingual advantage appears elusive, and seems to be related to both a specific sample of participants as well as specific tasks (Ross & Melinger, 2017). However, given that so many people are bi- and multilingual, understanding the cognitive effects of bi- and multilingualism, be they positive, negative, or simply nonexistent, becomes an important issue in trying to understand how a large part of the planet thinks and understands the world. The level of controversy within the subject should elicit a more focused investigation, centered on trying to understand exactly what sets of circumstances elicit a bilingual advantage
and why so many do not; a systematic investigation into these issues should lead us to an understanding of what, if any, the cognitive effects of multilingualism are.
References


Tables

**Table 1: Descriptives**

<table>
<thead>
<tr>
<th>All participants</th>
<th>Stroop congruent (ms)</th>
<th>Stroop incongruent (ms)</th>
<th>Stroop control (ms)</th>
<th>Flanker congruent (ms)</th>
<th>Flanker incongruent (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>62</td>
<td>62</td>
</tr>
<tr>
<td>Mean</td>
<td>948.23</td>
<td>1145.21</td>
<td>962.75</td>
<td>560.47</td>
<td>591.23</td>
</tr>
<tr>
<td>SE</td>
<td>36.94</td>
<td>46.88</td>
<td>35.91</td>
<td>14.48</td>
<td>14.22</td>
</tr>
</tbody>
</table>

By language group

<table>
<thead>
<tr>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monolinguals</td>
<td>-209.15</td>
<td>208.63</td>
<td>-280.86</td>
<td>362.20</td>
<td>-168.75</td>
<td>307.43</td>
</tr>
<tr>
<td>Bilinguals</td>
<td>40.06</td>
<td>208.94</td>
<td>60.29</td>
<td>210.70</td>
<td>-185.88</td>
<td>353.54</td>
</tr>
<tr>
<td>Multilinguals</td>
<td>-169.08</td>
<td>173.77</td>
<td>-220.57</td>
<td>331.95</td>
<td>-354.63</td>
<td>372.39</td>
</tr>
</tbody>
</table>

**EOWPVT Standardized Score**

<table>
<thead>
<tr>
<th>N</th>
<th>23</th>
<th>29</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>98.65</td>
<td>13.69</td>
<td>99.86</td>
<td>17.824</td>
</tr>
</tbody>
</table>

*Table 1: Stroop effect was calculated subtracting the incongruent condition reaction time (RT) from the congruent condition RT; a negative number indicates decreased speed on the incongruent condition. The flanker effect was calculated in the same way. The Stroop cost was calculated by subtracting the incongruent RT from the control condition RT; a negative number indicates decreased speed on the incongruent condition relative to the control. The Stroop facilitation was calculated by subtracting the congruent RT from the control RT; a positive number indicates that participants were faster on the congruent than control condition.*
Table 2: Monolinguals vs. bi/multilinguals

<table>
<thead>
<tr>
<th></th>
<th>Analyzed by</th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stroop effect</strong></td>
<td>Language Category</td>
<td>1, 56</td>
<td>.598</td>
<td>.443</td>
<td>.011</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Language Category *</td>
<td>1, 56</td>
<td>2.464</td>
<td>.122</td>
<td>.044</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Language Category</td>
<td>1, 56</td>
<td>.812</td>
<td>.372</td>
<td>.015</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stroop cost</strong></td>
<td>Language Category</td>
<td>1, 56</td>
<td>.051</td>
<td>.822</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Language Category</td>
<td>1, 56</td>
<td>.345</td>
<td>.560</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Language Category</td>
<td>1, 56</td>
<td>.005</td>
<td>.944</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stroop facilitation</strong></td>
<td>Language Category</td>
<td>1, 56</td>
<td>1.362</td>
<td>.248</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Language Category</td>
<td>1, 56</td>
<td>1.236</td>
<td>.271</td>
<td>.023</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Language Category</td>
<td>1, 56</td>
<td>1.271</td>
<td>.265</td>
<td>.023</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Flanker effect</strong></td>
<td>Language Category</td>
<td>1, 58</td>
<td>.001</td>
<td>.981</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>EOWPVT Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Language Category</td>
<td>1, 58</td>
<td>.465</td>
<td>.498</td>
<td>.008</td>
</tr>
<tr>
<td></td>
<td>EOWPVT Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Language Category *</td>
<td>1, 58</td>
<td>.006</td>
<td>.941</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Stroop effect, cost, and facilitation as well as flanker effect were calculated as described in Table 1. Language category refers to whether an individual was monolingual or bi/multilingual. EOWPVT score refers to the standardized rather than raw score, as the standardization provided a means to compare between the EOWPVT-4 and EOWPVT-4:SBE. No significant effects or interactions were found.
Table 3: Monolinguals vs. Bilinguals vs. Multilinguals

<table>
<thead>
<tr>
<th></th>
<th>Analyzed by</th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroop effect</td>
<td>Language Group</td>
<td>2, 56</td>
<td>.321</td>
<td>.727</td>
<td>.012</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td>1, 56</td>
<td>1.115</td>
<td>.296</td>
<td>.021</td>
</tr>
<tr>
<td></td>
<td>Language Group * EOWPVT score</td>
<td>2, 56</td>
<td>.436</td>
<td>.649</td>
<td>.017</td>
</tr>
<tr>
<td>Stroop cost</td>
<td>Language Group</td>
<td>2, 56</td>
<td>.069</td>
<td>.933</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td>1, 56</td>
<td>.504</td>
<td>.481</td>
<td>.010</td>
</tr>
<tr>
<td></td>
<td>Language Group * EOWPVT score</td>
<td>2, 56</td>
<td>.030</td>
<td>.970</td>
<td>.001</td>
</tr>
<tr>
<td>Stroop facilitation</td>
<td>Language Group</td>
<td>2, 56</td>
<td>.971</td>
<td>.386</td>
<td>.037</td>
</tr>
<tr>
<td></td>
<td>EOWPVT score</td>
<td>1, 56</td>
<td>.156</td>
<td>.694</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td>Language Group * EOWPVT score</td>
<td>2, 56</td>
<td>.843</td>
<td>.436</td>
<td>.032</td>
</tr>
<tr>
<td>Flanker effect</td>
<td>Language Group</td>
<td>2, 58</td>
<td>.215</td>
<td>.807</td>
<td>.008</td>
</tr>
<tr>
<td></td>
<td>EOWPVT Score</td>
<td>1, 58</td>
<td>.335</td>
<td>.565</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>Language Group * EOWPVT score</td>
<td>2, 58</td>
<td>.399</td>
<td>.673</td>
<td>.015</td>
</tr>
</tbody>
</table>

Table 3. Stroop effect, cost, and facilitation as well as flanker effect were calculated as described in Table 1. Language group refers to whether an individual was monolingual, bilingual, or multilingual. EOWPVT score refers to the standardized rather than raw score, as the standardization provided a means to compare between the EOWPVT-4 and EOWPVT-4:SBE. No significant effects or interactions were found.
Appendix

Language Experience and Proficiency Questionnaire (LEAP-Q)


Participant Code:

Study Code:

Age:

M/F:

(1) Please list ALL the languages you know in **order of dominance**:

(2) Please list ALL the languages you know in **order of acquisition** (your native language first):

(3) Please list what percentage of the time you are currently and on average exposed to each language. *(Your percentages should add up to 100%):*

<table>
<thead>
<tr>
<th>List language here:</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>List percentage here:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(4) When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you. *(Your percentages should add up to 100%):*

<table>
<thead>
<tr>
<th>List language here:</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>List percentage here:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(5) When choosing a language to speak with a person who is equally fluent in all your language, what percentage of time would you choose to speak each language? Please report percent of total time. *(Your percentages should add up to 100%):*

<table>
<thead>
<tr>
<th>List language here:</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
(6) Please name the cultures with which you identify. On a scale of zero to ten, please rate the extent to which you identify with each culture. (Examples of possible cultures include US-American, Chinese, Jewish-Orthodox, etc):

<table>
<thead>
<tr>
<th>List culture here:</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale (0-10) rating here:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(7) How many years of formal education do you have? ___

What is your highest education level? ____

(8) Have you ever had a vision problem, hearing impairment, language disability, or learning disability? (Select any that apply). If yes, please explain (including any corrections): ___

(9) Which of the following terms best describes your language proficiency? Monolingual (proficient in one language), bilingual (proficient in two languages), or multilingual (proficient in three or more languages)?

For each of the languages listed in questions (1) and (2) above, please complete the following (can be repeated as many times as necessary):

**Language name:**

**This is my** (first/second/third/fourth/etc) **language.** ____

(1) Age when you...:

<table>
<thead>
<tr>
<th>...began acquiring the language:</th>
<th>...became fluent in the language:</th>
<th>...began reading in the language:</th>
<th>...became fluent reading in the language:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(2) Please list the number of years and months you spent in each language environment:

<table>
<thead>
<tr>
<th>Years</th>
<th>Months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A country where the language is spoken.  
A family where the language is spoken.  
A school and/or working environment where the language is spoken.

(3) On a scale from zero to ten, please select your level of proficiency in speaking, understanding, and reading in this language. 0 = none, 5 = adequate, 10 = perfect
- Speaking: ___
- Understanding spoken language: ___
- Reading: ___

(4) On a scale from zero to ten, please select how much the following factors contributed to you learning the language. 0 = not a contributor, 5 = moderate contributor, 10 = most important contributor
- Interacting with friends: ___
- Interacting with family: ___
- Reading: ___
- Language tapes/self-instruction: ___
- Watching TV: ___
- Listening to the radio: ___
- Formal instruction (i.e. a class/course): ___

(5) Please rate on a scale of zero to ten to what extent you are currently exposed to the language in the following contexts. 0 = never, 1 = almost never, 5 = half of the time, 10 = always
- Interacting with friends: ___
- Interacting with family: ___
- Watching TV: ___
- Listening to radio/music: ___
- Reading: ___
- Language lab/self-instruction: ___

(6) In your perception, how much of a foreign accent do you have in the language?
0 – none
1 – almost none
2 – very light
3 – light
4 – some
5 – moderate
6 – considerable
7 – heavy
8 – very heavy
9 – extremely heavy
10 – pervasive

(7) Please rate how frequently others identify you as a non-native speaker based on your accent in the language. 0 = never, 1 = almost never, 5 = half the time, 10 = always: ______

(8) Do you consider yourself to be at least conversationally proficient? (i.e., are you able to participate in conversation with a native speaker comfortably?)