LSHSS

Research Article

Do Bilingual Children Have an Executive Function Advantage? Results From Inhibition, Shifting, and Updating Tasks

Genesis D. Arizmendi,[a] Mary Alt,[a] Shelley Gray,[b] Tiffany P. Hogan,[c] Samuel Green,[b,†] and Nelson Cowan[d]

Purpose: The purpose of this study was to examine differences in performance between monolingual and Spanish–English bilingual second graders (aged 7–9 years old) on executive function tasks assessing inhibition, shifting, and updating to contribute more evidence to the ongoing debate about a potential bilingual executive function advantage.
Method: One hundred sixty-seven monolingual English-speaking children and 80 Spanish–English bilingual children were administered 7 tasks on a touchscreen computer in the context of a pirate game. Bayesian statistics were used to determine if there were differences between the monolingual and bilingual groups. Additional analyses involving covariates of maternal level of education and nonverbal intelligence, and matching on these same variables, were also completed.
Results: Scaled-information Bayes factor scores more strongly favored the null hypothesis that there were no differences between the bilingual and monolingual groups on any of the executive function tasks. For 2 of the tasks, we found an advantage in favor of the monolingual group.
Conclusions: If there is a bilingual advantage in school-aged children, it is not robust across circumstances. We discuss potential factors that might counteract an actual advantage, including task reliability and environmental influences.

The possibility of a *bilingual cognitive advantage* has been suggested for decades (e.g., Bialystok & Martin, 2004; Diaz, 1985; Peal & Lambert, 1962). This refers to research findings demonstrating that bilinguals often outperform monolinguals on tasks that tap into executive functions such as those requiring inhibition, shifting, and updating. However, this idea of an advantage is contested. Over the past several years, there has been an increase in the number of studies that support these claims (e.g., Adesope, Lavin, Thompson, & Ungerleider, 2010) as well as those that refute them (e.g., de Bruin, Treccani, & Della Sala, 2015). Thus, the existing literature on whether there are differences between monolingual and bilingual individuals, particularly children, is mixed. This has ramifications for how we understand bilingual development. There are three general questions associated with this topic: (a) Is there a bilingual advantage for school-aged children? (b) If there is a bilingual advantage, is it restricted to certain types of executive functions? (c) What might explain some of the discrepant findings in the literature? This article will compare the performance of school-aged monolingual and bilingual children on a range of executive function tasks that cover all three domains of executive function (i.e., inhibition, shifting, and updating), and our findings will be discussed in the context of extent literature on a cognitive advantage in bilingual children.

Issues of Theory and Measurement

Executive Functions/The Central Executive

Let us begin with some clarification of terminology. Executive functions are "general-purpose control mechanisms that modulate the operation of various cognitive subprocesses and thereby regulate the dynamics of human

[a]Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson
[b]Arizona State University, Tempe
[c]MGH Institute of Health Professions, Boston, MA
[d]University of Missouri–Columbia

[†]In memory of our colleague and collaborator, Samuel (Sam) Green, who passed away during the preparation of this manuscript. We gratefully acknowledge his contributions to the research.

Correspondence to Genesis D. Arizmendi: genesis@email.arizona.edu

Editor-in-Chief: Sean Redmond
Editor: Ron Gillam
Received October 6, 2017
Revision received February 12, 2018
Accepted February 18, 2018
https://doi.org/10.1044/2018_LSHSS-17-0107
Publisher Note: This article is part of the Clinical Forum: Working Memory in School-Age Children.

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

cognition" (Miyake et al., 2000, p. 50). The term "executive function" is sometimes confused with the concept of the "central executive." The central executive is best known as part of the multicomponent model of working memory, a tri-component system used to describe the link between short-term and long-term memory (Baddeley & Hitch, 1974).[1] Thus, the "central executive" is a particular construct that can be used to analyze executive functions (Baddeley, 1998). We will be referring more generally to executive functions.

Conceptually, executive function could be viewed as any aspect of cognitive processing for which an individual has a choice. As such, executive functions are any aspect of processing that can be modified in a manner favorable to the individual when the individual is motivated to behave in a particular way. Jurado and Rosselli (2007) reviewed the concepts and components of executive functions defined by researchers over the years. Examples included volition, purposeful action, effective performance, concurrent manipulation of behavior, determination, planning, conscious actions, setting goals, strategy control and monitoring, abstract thinking, reasoning, inhibiting actions or behaviors, creative thinking, cognitive flexibility, problem-solving, organization, formation of concepts, and task analysis. This is a wide range of behaviors that, at first glance, can seem overwhelming and difficult to test. Happily, Miyake et al. (2000) used a latent variable analysis and narrowed executive functions down to three core functions, outlined in Table 1. We will use Miyake et al.'s terminology for the remainder of our discussion. Miyake et al. discussed how there is "unity and diversity" within executive functions. That is, there is enough in common with different executive function tasks and domains to justify grouping them under a single heading, yet there are enough differences that it might not be appropriate to conceptualize executive functions as a single concept. The take-home point for researchers is that conclusions made about executive functions should be precise in terms that describe which aspect of the "family of function" (Friedman, 2016, p. 541) one is referring to.

Potential Reasons for the Bilingual Cognitive Advantage in School-Aged Children

Before reviewing the evidence for a bilingual advantage, it would be useful to explain some potential reasons for an advantage. The bilingual advantage has been defined as bilinguals outperforming monolinguals on cognitive tasks tapping into executive functions in terms of improved accuracy, decreased reaction time, or both. Two primary types of explanations proposed to explain a bilingual advantage on cognitive tasks include domain-specific and domain-general explanations. The difference between these two types of explanations is whether the proposed bilingual advantage is restricted to tasks that draw upon inhibition (domain-specific) or whether bilingualism provides advantages to all core domains of executive function (domain-general).

[1]This construct was adapted from the supervisory attentional system (Norman & Shallice, 1986; Shallice, 1982).

Table 1. Core executive function domains and associated processes.

Core executive function domains (Miyake et al., 2000)	Processes associated with those functions
Inhibition	Inhibitory control: self-control, behavioral inhibition Interference control: selective attention, cognitive inhibition
Shifting	Set shifting, mental flexibility, mental set shifting, creativity
Updating	System for temporary storing, processing and manipulating information necessary for complex cognitive tasks

A *domain-specific* explanation is the Bilingual Inhibitory Control Advantage hypothesis, which states that bilinguals must frequently engage in inhibitory processes when selecting each of their languages, resulting in more efficient inhibitory processing (e.g., Green, 1998; Hilchey & Klein, 2011). Because of the continuous juggling of the two languages, the brain becomes more efficient at resolving conflicts resulting from the interference, or influence, of one language on the other (Bialystok, Craik, Klein, & Viswanathan, 2004). This hypothesis would predict that the bilingual advantage would emerge primarily in tasks that tax inhibition.

A *domain-general* hypothesis is the Bilingual Executive Processing Advantage (Hilchey & Klein, 2011), which predicts a bilingual advantage across any of the executive function domains. Examples of advantages in shifting tasks have been reported as early as 1962, when Peal and Lambert documented that bilingual French–English children were more accurate than their monolingual peers on nonverbal tasks that required symbolic "flexibility." They proposed that bilinguals may have demonstrated an advantage in this domain because people who learn two languages must learn two symbols for every object. Because of this, Peal and Lambert (1962) proposed that bilinguals become more efficient at concept formation and abstract thinking on tasks that required symbolic reorganization. Another hypothesis was that bilinguals have developed more flexibility in thinking. They stated that bilinguals have experience shifting between languages, particularly in cases where they may need to solve a problem. That is, bilinguals may attempt to think about a problem in one language but, if blocked, can "switch" to thinking about it in their other language. The ability to do this, whether conscious or unconscious, then may give bilinguals the ability to perform better on the tasks requiring symbolic organization. This is the "readiness to drop one hypothesis or concept and try another" ability (p. 14). Over the next 20 years, evidence continued to support the notion that bilingualism fosters a certain degree of "cognitive flexibility" (Diaz, 1985).

Sources of Difference in Findings in the Literature

There have been many reports that confirm these early findings of a cognitive advantage for bilinguals. However, there has also been some controversy about these results. In Table 2, we report studies in the literature that

Table 2. Summary of studies investigating executive function differences in monolingual and bilingual school-aged children.

Authors	Ages *N*	Languages	Country	Executive function components assessed (task name)	Number of indicators per component	Reliability of tasks reported	Results	Measure RT or ACC
Bialystok (1999)	5;0–6;3 *N* = 30 Mono = 15 Bi = 15	Mono = English Bi = English/Chinese	Canada	**Inhibition** Moving Word task	1	Not reported	MONO < BI	ACC *Inconsistent items only
				Shifting Dimensional Change Card Sort task	1		MONO < BI	ACC *Post-switch only
				Updating Visually Cued Recall task	1		MONO = BI	
Morton & Harper (2007)	6;0–7;0 *N* = 34 Mono = 17 Bi = 17	Mono = English Bi = English/French	Canada	**Inhibition** Simon task	1	Not reported	MONO = BI	
				Shifting	0			
				Updating	0			
Bialystok & Viswanthan (2009)	7;1–9;4 *N* = 90 Mono = 30 Bi = 30 Bi India =30	Mono = English Bi = English/Cantonese, Croatian, French, Hebrew, Hindi, Kannada, Mandarin, Marati, Punjabi, Russian, Tagalog, Telugu, Urdu Bi India = English and Tamil or Telugu	Canada and India	**Inhibition** Faces task	1	Not reported	MONO < BI	RT
				Shifting Faces task	1		MONO < BI	RT
				Updating Sequencing Span task	1		MONO = BI	
Carlson & Meltzoff (2008)	4;8–6;9 *N* = 50 Mono = 17 Bi = 12 Immersion = 21	Mono = English Bi = English/Spanish *Immersion = English and Spanish or Japanese *Native English when entered school with immersion program	USA	**Inhibition** Simon Says	5	Not reported	MONO = BI	
				Attention Network task			MONO = BI	
				Delay of Gratification			MONO = BI	
				Statue task			MONO = BI	
				Gift Delay With Cover			MONO = BI	
				Shifting Advanced DCCS task	1		MONO < BI	ACC
				Updating Visually Cued Recall	1		MONO < BI *Differences only emerge when controlling for age, SES, and verbal ability	ACC

(*table continues*)

Table 2. *(Continued).*

Authors	Ages *N*	Languages	Country	Executive function components assessed (task name)	Number of indicators per component	Reliability of tasks reported	Results	Measure RT or ACC
Bonifacci et al. (2011)	6;0–12;0 *N* = 36 Mono = 18 Bi = 18	Mono = Italian Bi = Italian and English, German, Chinese, Tagalog, Moroccan Arabic, Albanian, Polish, Slovack, Russian	Italy	**Inhibition** Go/No-Go task **Shifting** Anticipation **Updating** Memory with number Memory with symbol	2 1 2 	Not reported 	 MONO = BI MONO < BI MONO = BI MONO = BI	 RT/ ACC
Engel de Abreu (2011)	5;9–6;8 *N* = 44 Mono = 22 Bi = 22	Mono = Luxembourgish Bi = Luxembourgish and French, Spanish, German, Dutch, Portuguese, Czech, or Italian	Luxembourg	**Inhibition** **Shifting** **Updating** Counting Recall task Backward Digit Recall Digit Recall task	0 0 3 	3 reported .81 to .89 .80 to .85 .84 to .91	 MONO = BI MONO = BI MONO = BI	
Engel de Abreu et al. (2012)	8;1–8;2 *N* = 80 Mono = 40 Bi = 40	Mono = Portuguese Bi = Portuguese and Luxembourgish	Luxembourg and Portugal	**Inhibition** Sky Search Flanker task **Shifting** **Updating** Odd-One-Out Dot Matrix	2 0 2 	Not reported 	 MONO = BI MONO < BI MONO = BI MONO = BI	RT *did not use difference scores
Poarch & van Hell (2012)[a]	6;8–7;1 *N* = 75 Mono =20 Bi = 18 Second-language learners = 19 Triling = 18	Mono = German Bi = German/English Second-language learners = German/English	Germany	**Inhibition** Simon task **Shifting** **Updating**	1 0 0	Not reported 	 MONO < BI 	 ACC
Kapa & Colombo (2013)	5;8–14;11 *N* = 79 Mono = 22 Early Bi = 21 Late Bi = 36	Mono = English Early and Late Bi = English/Spanish	USA	**Inhibition** Attention Network Test **Shifting** **Updating** Forward Digit Span task	1 0 1 	Not reported 	MONO < EARLY BI MONO = BI	RT *did not use difference scores

(*table continues*)

Table 2. *(Continued).*

Authors	Ages *N*	Languages	Country	Executive function components assessed (task name)	Number of indicators per component	Reliability of tasks reported	Results	Measure RT or ACC
Morales, Calvo, & Bialystok (2013)	5;4–6;9 *N* = 56 Mono = 29 Bi = 27	Mono = English Bi = English and Arabic, Bulgarian, Cantonese, Chinese, French, Hebrew, Igbo, Mandarin, Portuguese, Russian, Serbian, Spanish, Urdu		**Inhibition** Pictures task, Conflict **Shifting** **Updating** Pictures task, Nonconflict Frog Matrices task	1 0 2 	Not reported	 MONO < BI MONO < BI MONO < BI	 RT/ACC RT ACC
Antón et al. (2014)	7;0–13;0 *N* = 360 Mono = 180 Bi = 180	Mono = Spanish Bi = Spanish/Basque	Spain	**Inhibition** Attention Network test **Shifting** **Updating**	1 0 0	Not reported	 MONO = BI 	
Duñabeitia et al. (2014)	8;0–12;0 *N* = 504 Mono = 252 Bi = 252	Mono = Spanish Bi = Spanish/Basque	Spain	**Inhibition** Classic Stroop Numerical Stroop **Shifting** **Updating**	2 0 0	Not reported	 MONO = BI MONO = BI 	
Filippi et al. (2015)	7;0–10;7 *N* = 40 Mono = 20 Bi = 20	Mono = English Bi = English and Italian, Spanish, Dutch, Armenian, Bengali, Polish, Czech, Russian, Portuguese	UK	**Inhibition** Sentence Interpretation Task **Shifting** **Updating** Forward Digit Span Backward Digit Span	1 0 2 	Not reported	 MONO < BI MONO = BI MONO = BI	ACC *only in one condition*
Antoniou et al. (2016)[b]	4;5–12;2 *N* = 136 Mono = 25 Multiling = 47 Bidialect = 64	Mono = Standard Modern Greek (SMG) Multiling = Greek, English, other Bidialect = Cypriot Greek and SMG	Greece	**Inhibition** Soccer task (Stop-Signal) Simon task **Shifting** Color Shape task **Updating** Backward Digit Span Corsi Blocks	1 1 1 	Not reported	MONO < BI	RT/ACC
Barac et al. (2016)	4;4–6;3 *N* = 62 Mono = 37 Bi = 25	Mono = English Bi = English and Spanish, French, Mandarin, Greek, Korean, Ukranian, Cantonese, Vietnamese, Tagalog, Russian, German, Polish	Canada	**Inhibition** Gift with Delay Attention Network task Go/No-Go task **Shifting** **Updating**	3 0 0	Not reported	 MONO = BI MONO < BI MONO < BI	ACC Congruous only RT/ACC

(*table continues*)

Table 2. *(Continued).*

Authors	Ages *N*	Languages	Country	Executive function components assessed (task name)	Number of indicators per component	Reliability of tasks reported	Results	Measure RT or ACC
Yang and Yang (2016)	5;0–6;0 *N* = 63 Mono = 31 Bi = 32	Mono = English Bi = Korean/English	USA	**Inhibition** Attention Network task	1	Test–Retest Reliability r = .94 for overall RT and r = .93 for error rate	MONO < BI	RT/ACC
				Shifting	0			
				Updating	0			
Blom et al. (2017)	6;0–7;9 *N* = 176 Mono = 44 Bi = 44 (3)	Mono = Dutch Bi = Frisian/Dutch Bi = Limburgish/Dutch Bi = Polish/Dutch	Netherlands	**Inhibition** Sky Search Flanker task	2	Not reported	 MONO < BI MONO = BI	RT[c]
				Shifting	0			
				Updating	2			
				Backward Digit Span			MONO = BI	
				Backward Dot Matrix			MONO = BI	
Ross & Melinger (2017)	6;0–9;0 *N* = 147 Mono = 45 Bi = 54 Bidialect = 48	Mono = English Bi = English and Gaelic, Arabic, Czech, Chinese, Malay, Russian, Japanese, Zulu, Greek, French	England and Scotland	**Inhibition** Simon task Flanker task	2	Not reported	 MONO < BI MONO = BI	 ACC
				Shifting	0			
				Updating	0			

Note. Mono = Monolingual; Bi = Bilingual; Triling = Trilingual; RT = reaction time; ACC = accuracy; DCCS = Dimensional Change Card Sort; SES = socioeconomic status.

[a]This study included additional measures, but those measures did not directly compare monolingual and bilingual groups. [b]This study clustered tasks to load onto principal component analysis. The Shifting task, loaded onto Inhibition, while the Updating tasks remained. Thus, due to the clustering of tasks in the analyses, all Inhibition tasks were counted as one task and the Updating tasks were counted as one task. [c]Used difference scores.

examine executive function differences between monolingual and bilingual school-aged children who have had at least one year of formal education (i.e., first grade and up). It is valuable to examine some of the sources of controversy for this literature to contextualize our findings. Additionally, insight into potential methodological pitfalls in the field may allow us to design future experiments that avoid them.

Unsurprisingly, given the overview of executive functions discussed above and the diversity of tasks that fall under the executive function heading, there is an enormous amount of variability in methodology across studies. These differences relate to the tasks selected, how performance is measured, and task categorization (e.g., inhibition vs. shifting). Few studies have explicitly examined all three domains of executive function across the same group of school-aged children, and none, to our knowledge, has used multiple indicators across all three domains. The implications of this methodological limitation go beyond the question of whether a putative advantage may be limited to a single executive function domain or may be a more general finding. There are additional issues with how to interpret the findings.

Friedman (2016) cautions against using a single measure of executive function to determine if an advantage exists because different executive function tasks often have low correlations with each other. We could liken this to the classic story of the blind men and the elephant, where each person's experience is unique (e.g., the trunk feels quite different from the hide, tusks, or ears), and to comprehend the animal, they have to combine their experiences. The implication is that, if we are to understand executive function, there are both conceptual and statistical reasons to ensure that we approach it comprehensively or that we tightly limit our interpretation based on the tasks that we use. For example, consider a study that only uses a Stroop task to measure executive function. This task only taps into inhibition, and thus, the researchers could not comment on updating and shifting and would need to modify their interpretations accordingly. In terms of statistics, we know that we need to use measures that load on the constructs that they aim to test. If we know that executive functions are best defined by multiple domains (e.g., Miyake et al., 2000), we cannot claim that a single measure represents the construct of executive functioning.

Another issue that muddies interpretation is the characteristics of the children who participate in the research studies. Bilingual individuals are a notoriously heterogeneous group. There can also be confounds with bilingualism and factors like socioeconomic status (SES) and culture, and researchers do not agree on the best ways to address these issues. For example, some researchers will use SES as a covariate (e.g., Antoniou, Grohmann, Kambanaros, & Katsos, 2016; Blom, Boerma, Bosma, Cornips, & Everaert, 2017; Carlson & Meltzoff, 2008; Chen, Zhou, Uchikoshi, & Bunge, 2014; Kapa & Colombo, 2013) to deal with the fact that there are real differences between the bilingual and monolingual populations and that matching would result in unrepresentative groups. However, this practice has been criticized by Paap, Johnson, and Sawi (2015), who pointed out that it may violate statistical assumptions to covary when the covariate and the groups are not independent, as is the case when the groups differ on the covariate measure. The alternative is matching, but Paap et al. (2015) noted that there are many alternative factors on which one might match, such as potential cultural, rather than SES, differences. While there is no clear, agreed-upon solution to these issues, it suggests that researchers need to consider these issues carefully and present an approach to minimize these confounding variables.

Aside from differences in culture and SES, it is important to consider the context in which a bilingual child is living. One possibility for differences in the literature is the linguistic environment of the children. Being bilingual has political, cultural, and sociolinguistic implications, all of which have the potential to support or mask a bilingual advantage. For example, a culture where bilingualism is the norm, and use of multiple languages is protected, provides a child with more opportunities to use both languages and a lack of stress related to "hiding" a perceived lower-status language. On the other hand, a culture that does not support bilingualism results in fewer opportunities for using both languages and has the potential to cause stress to a child who feels the need to limit the use of one language in certain contexts. This is especially important to consider in terms of whether bilingualism results in negative connotations, or if bilinguals feel the need to inhibit use of one language over another for fear of not fitting in with the majority population and/or discrimination. In the United States, racial–ethnic discrimination is pervasive (Telles & Murguia, 1990) and brings with it acculturation and acculturation stress. Acculturation is the process of cultural change that occurs when a person encounters another culture, leading to acculturation stress, which arises from the struggle to mesh the culture of origin to the host culture (Kulis, Marsiglia, & Nieri, 2009). Importantly, acculturation is often indicated by several related factors, including language proficiency, language use, nativity status, cultural-related behavioral preferences, and ethnic identity (Martinez, 2006). It would help with interpretation of the literature if researchers reported on the cultural and social contexts (beyond SES) in which the languages of the children are used.

The last, most troubling issues we discuss are the possibility of a publication bias (de Bruin, Treccani, & Della Sala, 2015) and the use of questionable statistical practices (Paap et al., 2015). de Bruin et al. compared conference abstracts and published works (including studies of both adults and children) that examined the bilingual advantage. They found far more studies published that supported the existence of an advantage compared to those in conference abstracts and were unable to attribute this difference to issues like sample size. However, some authors question the methodology of de Bruin et al.'s (2015) findings. Bialystok, Kroll, Green, MacWhinney, and Craik (2015) pointed out that conference findings are often different from submitted

articles in that they may consist of more preliminary data with smaller sample sizes and are often not subject to the same degree of peer review. Perhaps more convincing is Paap et al.'s (2015) meta-analysis that showed that the bilingual advantage only appeared in studies with lower sample sizes ($N < 50$), and not in studies with higher Ns. Statistically, this is not consistent with a robust effect and is suggestive of publication bias, although there is clearly room for debate. For example, does increasing sample size often come with increasing sample diversity that could mask an effect that could have been obtained in a homogeneous subsample?

Possible Bilingual Advantages in Executive Function in School-Aged Children

Inhibition

Inhibition is the domain of executive function that has been most well studied. Of the studies we found assessing executive function in school-aged bilinguals, only one did not include a measure of inhibition. Many studies investigating school-aged executive function differences opt to use similar tasks to measure this domain, including the Simon task, the Attention Network task, Go/No-Go, and the Flanker task. Few of these studies used difference scores as a dependent variable, which is important to do because it can otherwise be difficult to clearly interpret results. For example, if one only looks at the reaction time for incongruent trials, one cannot account for potential differences in the overall reaction time that might be driving the difference between groups. In other words, if a child is faster on all types of trials, it does not mean that there is something special about the incongruous trials. This is an example of the "task impurity" that Friedman (2016) referred to. None of these tasks report reliability.

Many executive function tasks provide two ways to find an advantage: accuracy and reaction time. In Table 2, there are more tasks with significant between-groups differences than tasks with no between-groups differences; however, most studies did not find an advantage for both accuracy and reaction time. The significant findings were roughly equally divided between advantages for accuracy or reaction time. Two studies that found a bilingual inhibition advantage used two measures, but in both cases, the differences emerged on only one task. The studies with significant between-groups differences were on the smaller side, with Ns < 50 per group, whereas the studies with no significant between-groups differences ranged from small Ns to Ns of over 250 per group. Thus, the available evidence is mixed in terms of whether or not the advantage appears and whether the advantage is in reaction time or accuracy. It would be useful to examine inhibition using multiple measures with larger Ns and measures to control for irrelevant variations in performance. An example of an irrelevant variation is overall reaction time. On any task, some individuals will likely have faster reaction times. However, overall reaction time does not provide insight into inhibition; it is irrelevant. Difference scores (between congruent and incongruent trials) are what illustrate the cost of the inhibitory response. Finally, it is important to use tasks that have strong reliability.

Shifting

Out of the three core domains of executive functions, there are few studies directly assessing shifting in school-aged children. The most commonly used measure among researchers is the Dimensional Change Card Sort task (e.g., Bialystok, 1999; Carlson & Meltzoff, 2008), though other studies have used other tasks (e.g., Color Shape task, Anticipation, Faces task). None of these studies reported reliability, nor did they take irrelevant variation into account, by using, for example, difference scores to capture the cost of shifting.

In our review of the literature for children who were at least 6 years old, we found four studies that assessed shifting. A bilingual advantage was found in all four, despite the fact that each study used a different task. Each of these studies was diverse in terms of country (i.e., Canada, India, United States, Italy), and each had a relatively small N (< 30). Three showed accuracy advantages and two showed reaction time advantages.

While the evidence points to differences in performance between monolingual and bilingual children in shifting, there are some lingering concerns about potential confounds. While it is encouraging to find these advantages across a range of tasks and children from different sociolinguistic backgrounds, a stronger case could be made if more studies assessed shifting using multiple reliable measures.

Updating

The third domain that we examined was updating. Ten of the 18 studies we reviewed analyzed updating using at least one indicator. There was variation in the characteristics of the children recruited, the country in which the study took place, and in the tasks that were used to assess updating. Some of the more commonly used measures included the Visually Cued Recall task and Forward and Backward Digit Span tasks. However, other tasks such as Odd-One-Out, Dot Matrix, Pictures task, and Frog Matrices were also used. What most of these studies had in common was the fact that they did not find a bilingual advantage. With the exception of Morales, Calvo, and Bialystok's (2013) study, who found bilingual advantages on one task in reaction time and one task in accuracy, all other studies reported that there were no differences between the groups. Carlson and Meltzoff (2008) originally reported that there were no differences between groups on their updating task; however, when they controlled for age, SES, and verbal ability, differences emerged.

Some studies used multiple indicators to assess updating. Only one study actually reported reliability for the tasks used (Engel de Abreu, 2011). The Ns for these studies were also relatively small (< 50). Taken together, the evidence does not seem to favor a bilingual advantage for updating. However, it would be helpful to replicate a study like Engel

de Abreu (2011) with a larger *N*. Larger sample sizes would help to control for a potential Type II error, in which a study would not reveal a true difference between groups solely due to an inadequate sample size.

Summary of Findings in the Literature

A review of research findings from the past 45 years resulted in studies that reported advantages and disadvantages across the three domains of the central executive. The results are difficult to interpret in terms of the Bilingual Inhibitory Control Advantage and Bilingual Executive Processing Advantage hypotheses. Clearly, if there is a domain-general advantage, it has not been robust enough to emerge across all tasks. However, findings for a bilingual advantage in executive function domains other than inhibition suggest that Bilingual Inhibitory Control Advantage may not be a comprehensive enough hypothesis.

The Present Study

The purpose of this study was to determine whether there were significant differences between monolingual and Spanish–English bilingual second-grade children on the executive function tasks of inhibition, shifting, and updating. By examining the three domains of executive functions, we were able to address the question whether any advantage is domain-specific or domain-general. This also allowed us to avoid methodological pitfalls related to a limited assessment of the construct. Additionally, we report reliabilities for the tasks used in the experiment.

This special issue focuses on the broad construct of working memory. Within that construct, our work focuses on the central executive component of working memory, including the three core executive functions of inhibition, shifting, and updating. In life, children must use each of these functions to accomplish tasks. If there are differences in executive function abilities between monolinguals and bilinguals, this provides valuable insight into potential differences in learning mechanisms and cognitive capacities that may be present in each population. Gaining additional insight into differences in cognitive processing allows us to revise theoretical models and assessments and to tailor interventions to best meet the needs of monolingual and bilingual children.

Method

Participants

One hundred sixty-seven monolingual and 80 Spanish–English bilingual children participated in this study.[2] Children were recruited through public schools in southern Arizona. After receiving institutional review board approval for the projects, parents were provided with information about the study printed in English on one side and in Spanish on the other. Parents who were interested in having their children participate returned the forms with their contact information to the school, and were then contacted by the research team. To determine if a child was monolingual or bilingual, we collected a detailed parent questionnaire about each child's linguistic environment. To qualify for the monolingual group, parents had to report that their child's primary language was English, that the primary caregivers for their child spoke English only, and that all prior and current academic instruction was in English only. To qualify for the bilingual group, parents had to report that their child could carry on a conversation in English and Spanish. Bilingual children could have either English or Spanish reported as their primary language. At least one primary caregiver needed to report speaking Spanish in the home to the child. All prior and current academic instruction could have included English, Spanish, or both (see Table 3 for qualifying measures).

All of the bilingual children had to complete the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003) and the Spanish Formulación de Oraciones subtest of the CELF-4 Spanish Edition (Semel, Wiig, & Secord, 2006). If the bilingual children did not obtain a standard score greater than or equal to 88 on the English CELF-4 (indicating they did not have language impairment), they had to complete the full Spanish CELF-4. Children who earned a standard score of 78 or better on the full Spanish CELF-4 were considered not to have language impairment. This score is empirically derived and has a sensitivity of 86% and a specificity of 80% for this population (Barragan, Castilla-Earls, Martinez-Nieto, Restrepo, & Gray, 2018). To confirm that they had sufficient proficiency in each language to form complete sentences, children had to earn a standard score of 6 or greater on both the English Formulated

Table 3. Inclusionary criteria to be classified as typically developing.

Measure	Criteria for all children
Vision Acuity	Pass screening
Color Vision	Pass screening
Hearing	Pass screening
Nonverbal Cognition (K-ABC2)	Standard Score ≥ 75
Word Reading (TOWRE-2)	Standard Score ≥ 96
Oral Language (CELF-4)[a]	Standard Score ≥ 88
Speech Skills (GFTA-2)	≥ 31st Percentile

Note. K-ABC2 = Kaufman Assessment Battery for Children–Second Edition (Kaufman & Kaufman, 2004); TOWRE-2 = Test of Word Reading Efficiency–Second Edition (Torgesen, Wagner, & Rashotte, 2012); CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel, Wiig, & Secord, 2003); GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000).

[a]Monolingual only.

[2]These children were part of a larger study: POWWER–Profiles of Word Learning and Working Memory for Educational Research, which was funded by National Institute on Deafness and Other Communication Disorders Grant R01 DC010784. The full working memory battery was described in Cabbage et al. (2017). Data from the typically developing participants have been reported on in Alt et al. (2017), Cowan et al. (2017), Gray et al. (2017), and Green et al. (2016).

Sentences subtest of the CELF-4 and the Spanish Formulación de Oraciones subtest of the CELF-4 Spanish Edition.

We also collected information about children's vocabulary from the Expressive Vocabulary Test–Second Edition (EVT-2; Williams, 2007), reading comprehension from the Woodcock Reading Mastery Test, Paragraph Comprehension Subtest (WRMT; Woodcock, 2011), and a parent rating scale on attention and behavior. In addition, we collected additional information on Spanish vocabulary using the Expressive One-Word Picture Vocabulary Test–Bilingual Version (EOWPVT; Brownell, 2001). Table 4 includes descriptive statistics for both groups.

A large number of children did not qualify for participation in the study for the following reasons: 112 were bilingual, but did not fit the definition of typically developing due to reports of parent/teacher concerns, history of or current enrollment in speech language or special education services, or had a medical diagnosis (e.g., attention-deficit/hyperactivity disorder [ADHD] or seizures); 89 did not meet the inclusionary criteria for the Test of Word Reading Efficiency–Second Edition (TOWRE-2; Torgesen, Wagner, & Rashotte, 2012); 81 were exposed to Spanish, but could not carry on a conversation in Spanish; 19 did not meet the inclusionary criteria for either the English or Spanish CELF-4 scores; 15 had reportedly repeated a grade; 14 could not carry on a conversation in English; 15 were bilingual, but were exposed to languages other than Spanish or English; seven failed the hearing screening; four failed the vision screening; and one child was reported to be bilingual, but none of the primary caregivers spoke Spanish.

Table 4. Means and standard deviations for standard scores on inclusionary and descriptive assessments.

Variable	Monolingual	Bilingual	*p* value[a]
N	167	80	
Age	7;7 (0;4)	7;9 (0;5)	.001
MLE	15.38 (1.65)	12.58 (2.56)	< .001
TOWRE-2	109.44 (8.40)	108.10 (7.75)	.227
K-ABC2	117.60 (15.52)	106.61 (11.77)	< .001
CELF-4	108.75 (9.58)	93.45 (9.10)	< .001
GFTA-2[b]	50.89 (8.53)	44.80 (10.67)	< .001
EVT-2	112.38 (10.95)	93.88 (8.88)	< .001
WRMT	108.22 (9.85)	102.40 (9.10)	< .001
ADHD	10.19 (8.76)	7.90 (7.99)	.065
SCELF-4 total		93.48 (11.81)	
SCELF-4-FO		10.74 (2.37)	
EOWPVT		110.15 (13.87)	
EOWPVT ratio		0.52 (0.22)	

Note. MLE = maternal level of education; TOWRE-2 = Test of Word Reading Efficiency–Second Edition (Torgesen, Wagner, & Rashotte, 2012); K-ABC2 = Kaufman Assessment Battery for Children–Second Edition (Kaufman & Kaufman, 2004); CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel, Wiig, & Secord, 2003); GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000); EVT-2 = Expressive Vocabulary Test–Second Edition (Williams, 2007); WRMT = Woodcock Reading Mastery Test, Paragraph Comprehension Subtest (Woodcock, 2011); ADHD = parental rating of ADHD behaviors using the ADHD Rating Scale–IV Home Version (DuPaul et al., 1998; lower scores on this measure reflect fewer concerns); SCELF-4 total = Spanish Clinical Evaluation of Language Fundamentals–Fourth Edition Standard Score (Semel et al., 2006); SCELF-4-FO = Spanish Clinical Evaluation of Language Fundamentals–Fourth Edition Formulación Oraciones Standard Score; EOWPVT = Expressive One-Word Picture Vocabulary Test–Bilingual Version Standard Scores (Brownell, 2001); EOWPVT ratio = Expressive One-Word Picture Vocabulary Test–Bilingual Version total raw Spanish words produced/total raw words produced.

[a]Between-groups differences were tested using *t* tests. [b]Percentile, rather than standard score.

Design and Stimuli

All children completed the executive function tasks as part of a larger battery of working memory and word learning tasks (not reported on here) administered on a touchscreen computer (see Cabbage et al., 2017, for an overview of working memory tasks and Alt et al., 2017, for an overview of word learning tasks). Games were administered in the context of a pirate adventure in which children could earn virtual coins for correct answers. These could then be redeemed at the virtual pirate store. The tasks were designed to assess inhibition, shifting, or updating (see Table 5). Children were required to pass a training block to proceed with a task. This ensured that only children who demonstrated understanding of a task contributed data. Each task took approximately 10 minutes to complete, and children completed the set of tasks across the span of five separate days. Importantly, all the tasks were designed to have low linguistic requirements to avoid a confound of language and executive function abilities (Friedman, 2016). The working memory tasks did not have language associated with them, other than the instructions and trainings that were initially presented to the children. These instructions were supplemented with visual information so that even children with language impairments could understand them. Children needed to demonstrate their understanding by passing a training. Following the training, no language is presented during the experimental tasks with the exception of well-known words like the number words in digit span tasks and the color words in the Stroop tasks. The only verbal language children are required to produce are either colors or numbers for three of the tasks.

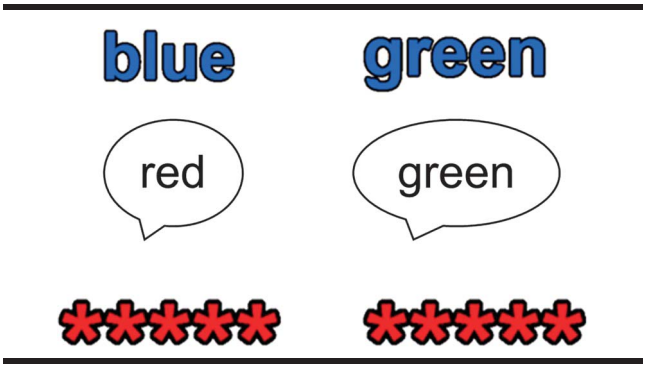
Inhibition Tasks

Two Stroop tasks were based on the classic task described by Stroop (1935). For our Classic Stroop task, children were asked to respond to congruous and

Table 5. Executive function tasks by domain.

Inhibition	Shifting	Updating
Classic Stroop	Pirate Sorting	Number Updating
Stroop Cross-Modal	Global Local	*N*-Back Auditory
Stop-Signal		*N*-Back Visual

Figure 1. Examples of Stroop Classic congruent (top left) and incongruent (top right) stimuli and Stroop Cross-Modal congruent (bottom left) and incongruent (bottom right) stimuli.



incongruous stimuli. For congruous stimuli, children saw a written word in a font color that matched the color of the word itself and were asked to either read the word ("read" block) or name the color of the font ("color" block). In the incongruous condition, children were asked to do the same thing, but the color of the word was incongruent with the color spelled out by the word (see Figure 1 for an example). For each block, there were 12 congruous and 12 incongruous items that were presented in a random order. Children responded verbally, and a research assistant entered their responses into the computer using color-coded buttons. Reaction times were scored by hand using audio files and Praat software (Boersma & Weenink, 2014). The dependent variable was the difference between the incongruous reaction time and the congruous reaction time. Only accurate responses were included. The Stroop Cross-Modal task followed the same logic and format of the Classic Stroop task, but the congruity/incongruity was between the color of the font of a series of asterisks and a recorded color word the child heard.

For the Stop-Signal task, children saw different "monsters" that were taking over an island. However, they were instructed that the "monsters" often looked like the "special pets" that lived on that island. The children were instructed to press the space bar if they saw a monster. One monster was presented at a time. If they heard the sound of a horn when they saw a monster, they were instructed that they should not press the space bar because that meant it was a "special pet" (see Figure 2). There were three experimental blocks that differed in terms of the time between the stimuli: simultaneous audio/visual, delay of 100 ms, and delay of 200 ms. There were 24 trials in each block, with a ratio of 1 stop:3 go trials. The dependent variable was corrected accuracy (GO – (1 –STOP)).

Shifting Tasks

For the Pirate Sorting task, children saw four different boats on the screen. Children were instructed to put the sea monster in the correct boat according to the instructions provided. The sea monsters could be sorted by color or shape. Each sea monster had a shirt that was either pink or yellow, with either circles or squares on the shirt. There were two color boats (yellow and pink) and two shapes of boats (circle or square), which were identified by their sails. In the middle of the screen, a flag changed to indicate which way the sea monsters should be sorted (e.g., a colorful rainbow when sorting by color; a black-and-white flag with shapes when sorting by shape). Children were to select the correct boat for each sea monster according to the flag. They made their selections by touching the boat on the touchscreen (see Figure 3). There were 32 trials of simple sorting (i.e., there were only two boats available for sorting) and 32 trials of complex sorting (i.e., there were four boats available for sorting). The dependent variable was the reaction time for the switch trials minus the reaction time for the same trials for accurate responses on the complex task.

Figure 2. Example of Stop Signal trials with a "GO" trial on the left and a "STOP" trial on the right.

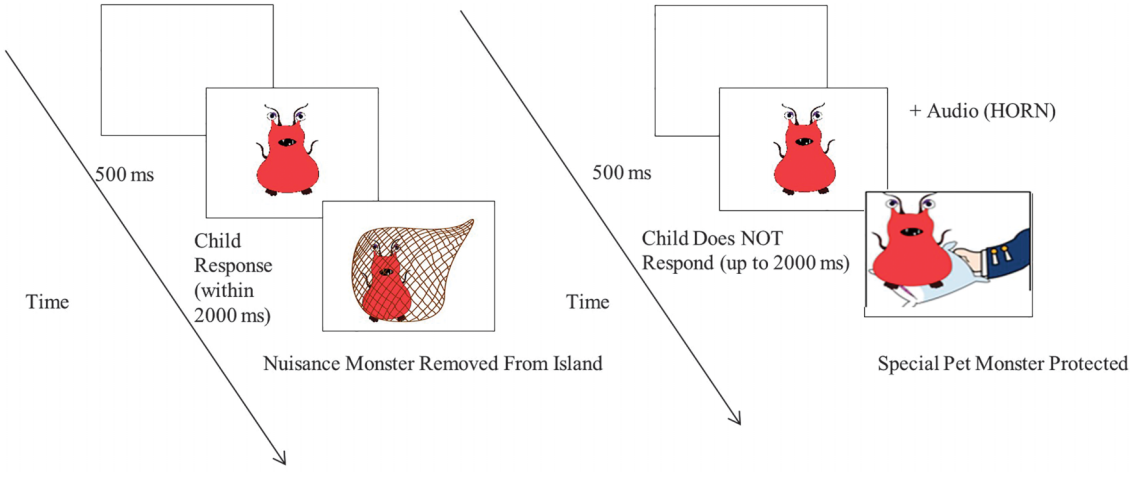
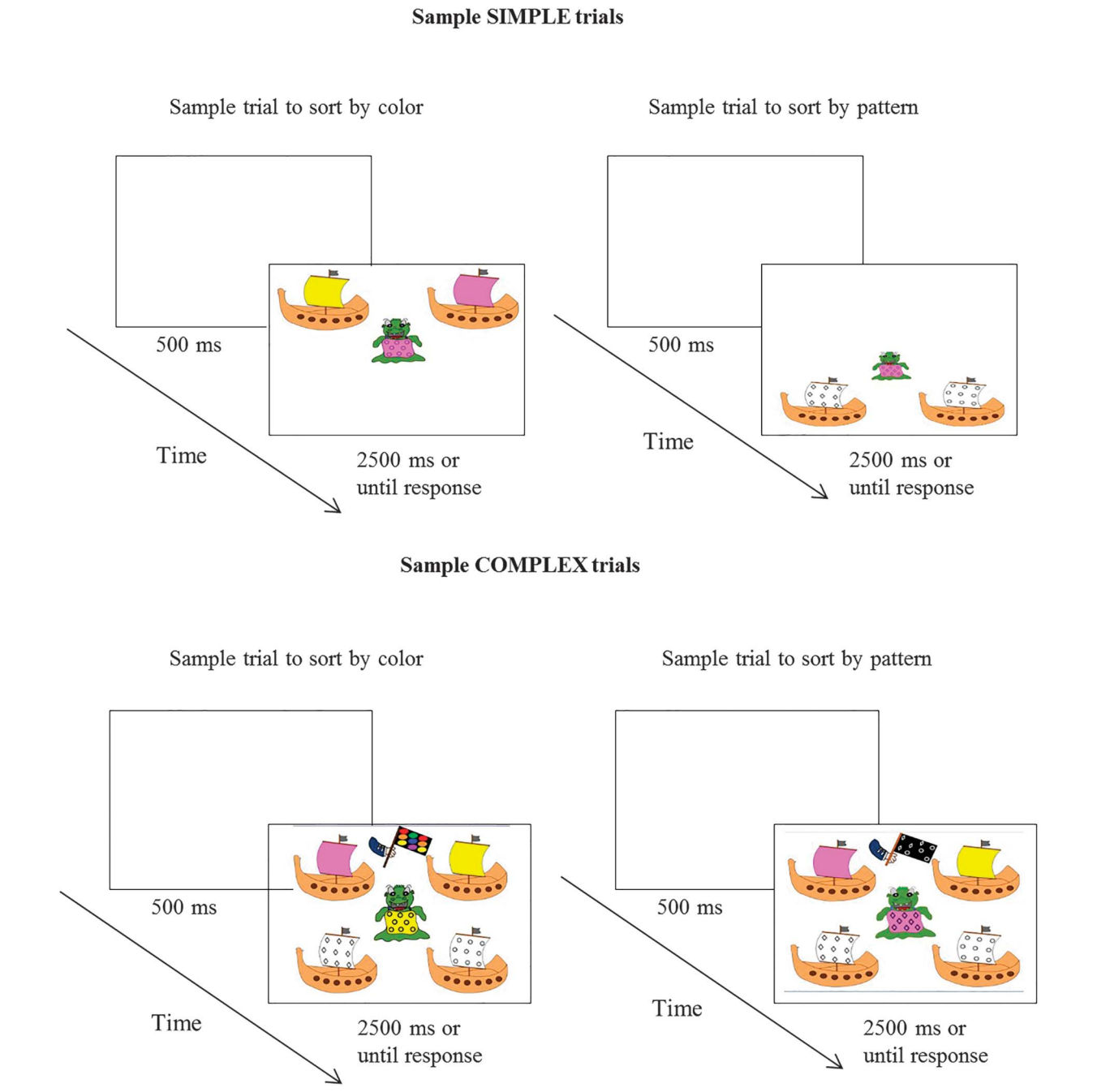
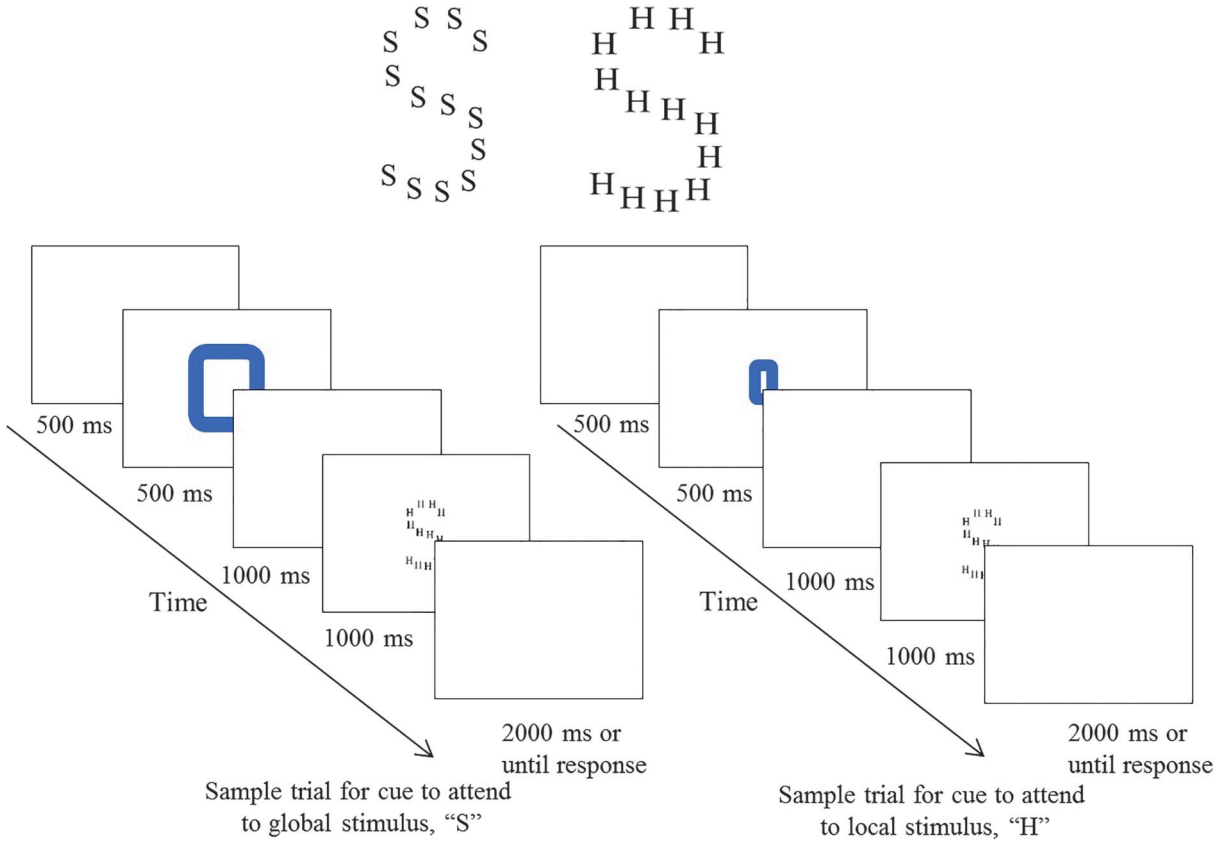


Figure 3. Pirate Sorting sequences for SIMPLE and COMPLEX trials. For the simple trials: (a) a cue to sort by color (shown in sequence on the left) and (b) a cue to sort by pattern (shown in sequence on the right). For the complex trials: (a) a cue to sort by color (shown in sequence on the left) and (b) a cue to sort by pattern (shown in sequence on the right).



The Global Local task required children to choose between the letters "H" and "S." This task was based on the one first described by Navon (1977). Children saw either an "H" or an "S." The large letter was composed of smaller letters that were either congruous (e.g., S made of tiny Ss) or incongruous (e.g., S made of tiny Hs). Children were trained to press a button marked "H" or "S" on the keyboard on the basis of a rectangle. If they saw a large rectangle, they had to select the large letter. If they saw a small rectangle, they had to select the small letter (see Figure 4). There was a single block with 24 trials evenly divided between global (large) and local (small) trials, which were also divided

Figure 4. Global Local; Sample stimuli for congruent (left) and incongruent stimuli (right), followed by task sequence for global (left) and local (right) trials.



into same versus switch trials. The dependent variable was the number of correct responses for same versus switch tasks.

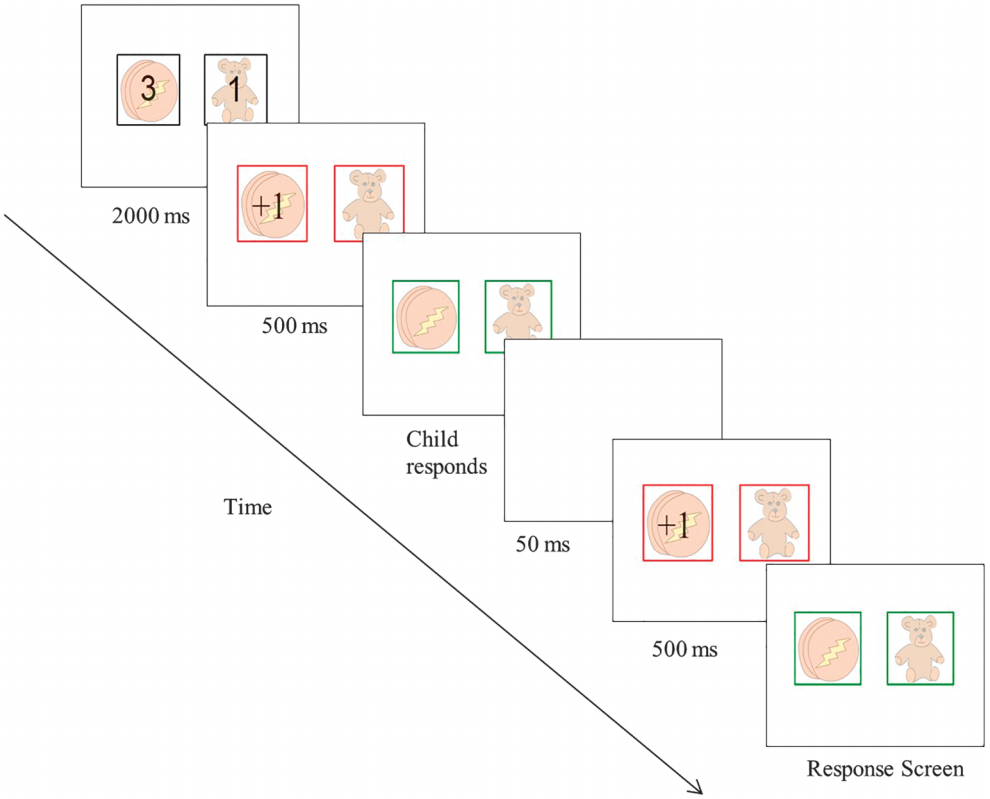
Updating

The Number Updating task was based on one described by Oberauer (2002). Our Number Updating task was in the context of a toy factory in which baby sea monsters wanted either a yo-yo or a teddy bear. Children would be presented with two numbers superimposed on images of a yo-yo and teddy bear. For example, to begin a block, children might see the number 1 on top of the yo-yos and a 3 on top of the teddy bears. Children were expected to remember those two numbers. The numbers disappeared, then children saw a 1 pop up under one of the toys. That number disappeared and children were instructed to add 1 to the appropriate toy and say the new number of yo-yos or teddy bears out loud. The number to be added could appear on either toy and children were expected to update the new numbers accordingly (see Figure 5). Children had three blocks of this game in which they made five updates to the toy order, resulting in a total of 15 updates. The dependent variable was lenient accuracy. That is, we allowed for the fact that an error early on could potentially lead to multiple errors, because all subsequent responses would be adding to the wrong base numbers. For example, if a child incorrectly responded with "2, 2" in the previous example and then was asked to update the right column, his or her successful response of "2, 3" would be incorrect, because it should have been "2, 4" by that point. By giving credit for correct updates, based on whatever the previous response was, we were sure not to underestimate performance.

Both *N*-Back tasks were based upon the one first described by Kirchner (1958). In the *N*-Back Auditory task, children listened to a tone and, 1000 ms later, heard another tone. Children were instructed to decide whether it was the same or different as the one heard directly before it by selecting a "green" key for same and a "red" key for different. This sequence continued with different tones being presented. Children only judged one tone back for all trials (see Figure 6). This game contained three blocks of 18 trials of which half contained "same" tones and half "different" tones, resulting in 54 data points total. The dependent variable was overall accuracy.

For the *N*-Back Visual task, children were presented with an image of a square with white dots inside of it for 1000 ms. The organization of the dots in the squares varied with each presentation. Children were instructed to decide

Figure 5. Example of the Number Updating task.



whether the square was the same or different as the one directly before it by selecting a "green" key for same and a "red" key for different. This sequence continued with different arrangements of the white dots in the square presented (see Figure 7). Like the *N*-Back Auditory task, children only recalled one image back for all trials. The number of trials and the dependent variable was the same as in the *N*-Back Auditory task.

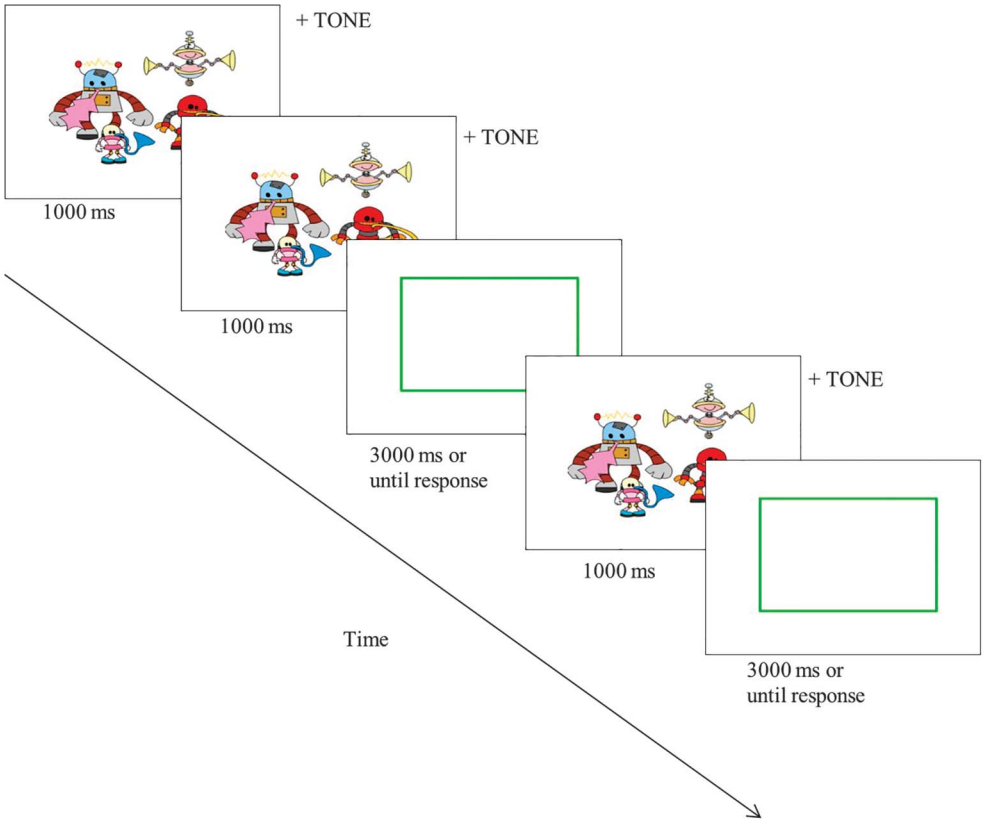
Analyses

Before analyzing group task performance, we examined the correlations among tasks. We also calculated each task's reliability using internal consistency coefficients as described by Green et al. (2016) using data from the monolingual group. Although there was a range of reliabilities (see Table 6), the inhibition tasks' reliabilities were below the level of .70, which is a minimum suggested threshold for basic research (Lance, Butts, & Michels, 2006 interpreting Nunnally, 1978). Therefore, we conducted no further analyses on these tasks.

Researchers examining bilingual populations need to make decisions about how to deal with the socioeconomic differences that are often present between bilingual and monolingual groups, especially in the United States. While it is possible to find children who match on these variables, this can lead to unrepresentative samples. Also, while SES has undeniable effects on early language outcomes, its role in second-grade children's language is less clear. Alt, Arizmendi, and DiLallo (2016) found SES to be a weak predictor of narrative language skills in English for second-grade bilingual children and found it to have no predictive abilities for these same children's Spanish language narratives. Nevertheless, to be safe, we explored our results using two different methods to account for the group differences in SES and nonverbal intelligence, our two measureable factors that were theoretically most likely to affect performance. First, we treated each factor as a covariate. Next, we created individual pairs matched on each of these factors. There were no substantial differences in the results using either of these techniques. The results of these specific analyses are reported in the Appendix. Thus, reported results include our full data set without using any covariates or matching.

To test for differences in between-groups performance, we utilized a Bayesian independent-samples *t* test (Rouder, Speckman, Sun, Morey, & Iverson, 2009). With this approach, a Bayes factor is calculated, which gives the probability that the data were in favor of the alternative hypothesis (i.e., a mean difference between groups) in comparison with the null hypothesis (i.e., no mean difference between the groups). Thus, the Bayes factor allows one to make a judgment about the merits of the alternative hypothesis relative to the null hypothesis. We conducted

Figure 6. Example of the *N*-Back Auditory task.



the Bayesian analyses using the Bayes factor calculator available at http://pcl.missouri.edu/bf-two-sample. We conducted these analyses using the scaled-information Bayes factor with a scale *r* on the effect size of .707.

The dependent variable for each task was chosen based on the nature of the task and the way that performance on these tasks is typically measured in the literature. We analyzed accuracy for the Number Updating and *N*-Back tasks (see Figure 8). For these tasks, a higher score indicates better performance. For these tasks, accuracy is the more telling variable; there is only one condition, and the issue is not how quickly one can update but how well one can do it. We analyzed the difference in reaction times on congruous and incongruous stimuli for Pirate Sorting (see Figure 9). For this task, a higher score is indicative of a greater cost associated with the incongruous task and thus reflects lower performance. The difference in reaction time is a more accurate measure of switching cost that can sometimes be hidden when looking only at accuracy.

Results

Group Differences

Applying the Bayesian independent-samples *t* test (Rouder et al., 2009), there was support in favor of the alternative hypothesis for the *N*-Back Visual task and the *N*-Back Auditory tasks favoring the monolingual group (see Table 7). The estimated Bayes factor for the *N*-Back Auditory tasks suggested that the data were 3.72 to 1 in favor of the alternative hypothesis in comparison with the null hypothesis. On the other five tasks, the results indicated various degrees of support for the null hypothesis (i.e., anecdotal to substantial). Overall, the findings of the Bayesian approach failed to support the bilingual advantage hypothesis.

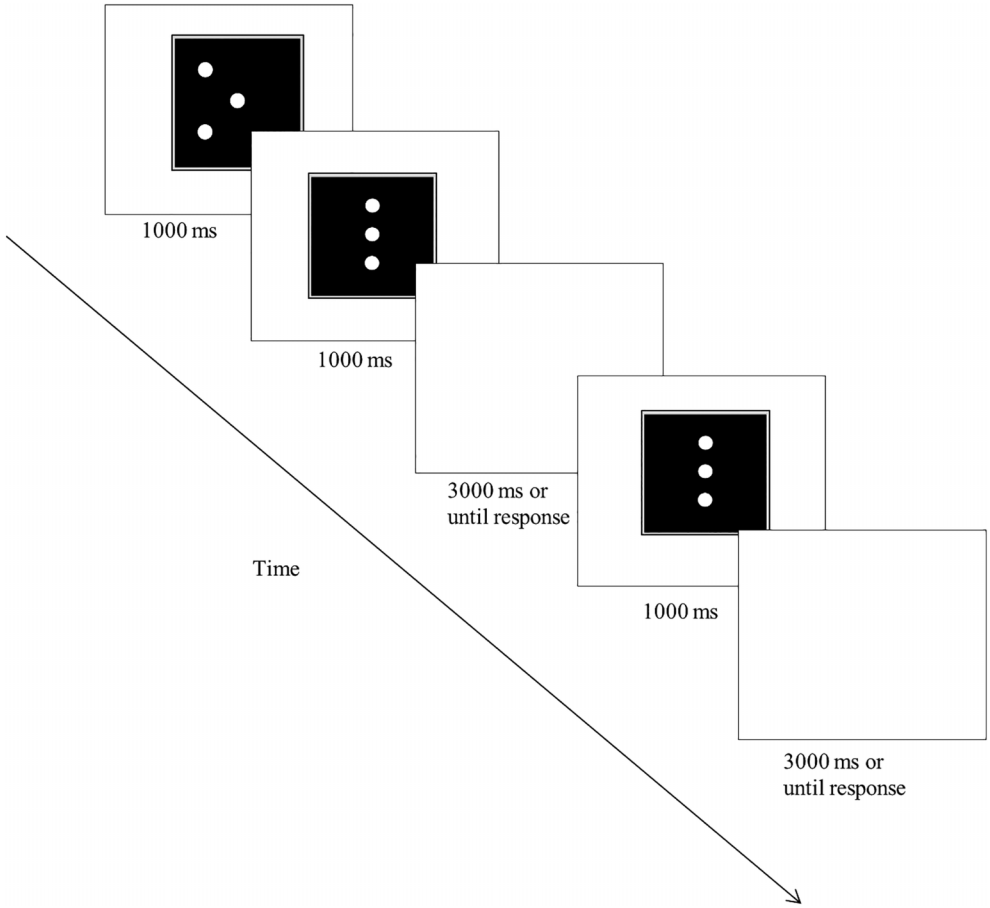
Correlations

We examined the correlations between our different indicators and domains using Pearson product correlations (see Table 8). The tasks were not all highly correlated with one another. Some, like the shifting task, were not correlated with any other tasks. However, we did see significant within-domain correlations for all of the updating tasks.

Discussion

Our work examined the performance between monolingual and Spanish–English bilingual second-grade children on tasks assessing executive functions, including inhibition,

Figure 7. Example of the *N*-Back Visual task.



updating, and shifting. Our data favored the null hypothesis. That is, it was more likely that there was not an observable bilingual executive function advantage in any domain we measured for bilingual second-grade students. In fact, the results of both *N*-Back tasks were strongly against the bilingual advantage hypothesis, favoring the monolingual group. Thus, this renders the point about the nature of the advantage moot. To consider reasons that some researchers find executive function advantages when we did not, we next consider the effects of differing bilingual environments and task reliability.

Bilingual Environments

As noted above, cultural context has the potential to impact outcomes. Below, we discuss two components of a cultural context that have the potential to explain why we may not have found evidence for between-groups differences.

Opportunities for Shifting

All of the bilingual children in our study came from southern Arizona. In their communities, the percentage of households who speak Spanish was 23.0% for one city and 76.2% for another (U.S. Census Bureau, 2015, 2011–2015 American Community Survey 5-Year Estimates).[3] In southern Arizona, children may not experience enrichment in their home language outside of the home. Specifically, all of the children in our study attended schools where English was the only language of instruction. At first glance, a potential consequence of this type of educational setting might be that bilingual children have fewer opportunities to shift between languages.

One potential interpretation of our findings is that our children simply did not have enough practice with shifting between languages to demonstrate a bilingual advantage. However, this explanation has limitations. First, it is not accurate to assume that bilingual individuals do not switch between languages just because the home and school languages are separate. This might be the case if a child's home language was different from all others in the community (e.g., a speaker of Hungarian in Arizona),

[3]We compared performance across both communities, but there were no significant differences.

Table 6. Internal consistency reliabilities by task.

Type of task	*N*	Reliability	95% CI
Classic Stroop			
Color: $RT_{Incongruent} - RT_{Congruent}$	156	.43	[.22, .58]
Color: $RT_{Incongruent}$		.75	[.66, .82]
Color: $RT_{Congruent}$		.62	[.48, .72]
Read: $RT_{Incongruent} - RT_{Congruent}$	156	.20	[.00, .42]
Read: $RT_{Incongruent}$		.88	[.84, .91]
Read: $RT_{Congruent}$		.69	[.57, .77]
Cross-Modal Stroop			
Color: $RT_{Incongruent} - RT_{Congruent}$	157	.41	[.19, .57]
Color: $RT_{Incongruent}$		.83	[.77, .88]
Color: $RT_{Congruent}$		.82	[.75, .87]
Repeat: $RT_{Incongruent} - RT_{Congruent}$	157	.20	[.00, .42]
Repeat: $RT_{Incongruent}$		.88	[.84, .91]
Repeat: $RT_{Congruent}$		.91	[.88, .93]
Pirate Sorting			
Simple: $RT_{Different} - RT_{Same}$	162	.82	[.75, .87]
Simple: $RT_{Different}$		.97	[.96, .98]
Simple: RT_{Same}		.95	[.93, .96]
Complex: $RT_{Different} - RT_{Same}$	162	.74	[.65, .81]
Complex: $RT_{Different}$		.88	[.84, .91]
Complex: RT_{Same}		.83	[.77, .88]
Global Local			
Local: $ACC_{Same} - ACC_{Different}$	136	.00	[.00, .29]
Local: $ACC_{Different}$		.33	[.06, .52]
Local: ACC_{Same}		.55	[.37, .68]
Global: $ACC_{Same} - ACC_{Different}$	136	.15	[.00, .39]
Global: $ACC_{Different}$		.54	[.35, .67]
Global: ACC_{Same}		.47	[.26, .62]
Stop Signal			
$ACC_{Go} - ACC_{No\ Go}$	158	.62	[.48, .72]
$ACC_{No\ Go}$		.66	[.53, .75]
ACC_{Go}		.88	[.84, .91]
Number Updating: Accuracy	139	.95	[.93, .96]
N-Back Visual: Accuracy	148	.86	[.81, .90]
N-Back Auditory: Accuracy	151	.82	[.75, .87]

Note. CI = confidence interval; RT = reaction time; ACC = accuracy.

but is not the case when the school language is English. The home language of the majority of the students is Spanish. In this case, you have a large community of children who are being taught in English, but could easily switch to Spanish when speaking with peers, at lunch, or in between lessons. We could also think about switching between languages as being either verbal or internal. Because a child may be listening to school instruction in English does not mean that they may not be processing material or thinking through information in Spanish. As is the case for many bilinguals, a language may be "shut off" to speak with different conversational partners, but that does not remove the fact that the second language plays an influence in the way a bilingual processes language (e.g., cross-linguistic influences). Also, the children in our study demonstrated strong skills in both of their languages. To be included in the bilingual sample, they were required to demonstrate proficiency in both languages. All children had the ability to easily converse in both languages using age-appropriate form, content, and use based on their use of grammar, semantics, and use of language, on the Formulated Sentences portion of the English and Spanish CELF-4. Some children could not meet our stringent inclusionary criteria in this regard. Therefore, if children with strong skills in both languages did not demonstrate a bilingual advantage, this suggests that children would need constant opportunities to verbally shift between languages to demonstrate an executive function bilingual advantage.

Another problem with this hypothesis is that even though there are not as many opportunities for verbal shifting in an English-only classroom, such a situation might actually increase the amount of inhibition that is needed. Recall that for most of our children, Spanish was their native language. Thus, in an English-only classroom, the need to suppress Spanish for verbal interactions would be increased. This should lead to a bilingual advantage in inhibition, which our findings did not support. So, while a lack of opportunity for verbal shifting does has the potential to impact executive function performance, it does not seem to be a satisfying answer with our particular children.

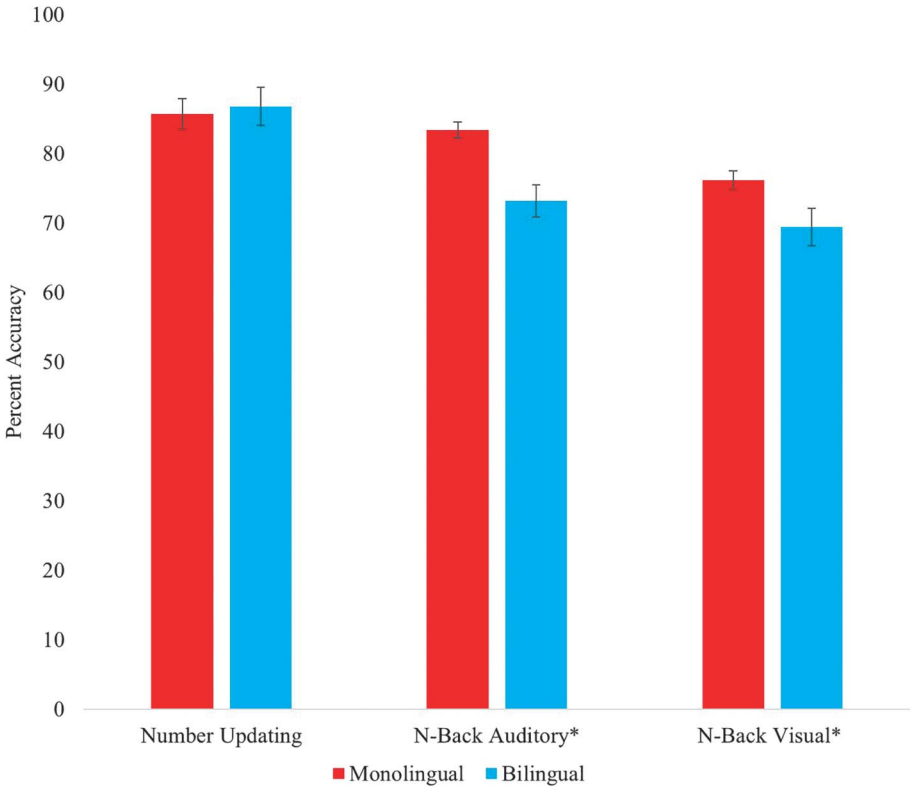
Acculturation and Acculturation Stress

When children only receive schooling in English, this is considered as a "subtractive" environment (Lambert, 1973). There is stress that comes from living in a subtractive environment. It is well documented that stress impairs executive functions (e.g., Arnsten, 1998; Blair, Granger, & Peters Razza, 2005; Diamond & Lee, 2011; Pechtel & Pizzagalli, 2011). It may be possible that this stress could counteract a potential bilingual executive function advantage.

What do we mean by stress associated with language? Many monolingual speakers think of bilingualism as being purely linguistic. That is, their consideration of bilingualism may be limited to which language a child chooses to use, or thinking about potential cross-linguistic influences that might affect how a bilingual child's language develops. However, language is entwined with cultural and social identities (e.g., Mahadi & Jafari, 2012). Choosing to use one language versus another can have social consequences for a bilingual speaker. For example, in states with large Spanish-speaking populations like California, Arizona, and Texas, a bilingual speaker at a store has the choice to speak Spanish or English. If the speaker chooses to use Spanish, she may encounter social obstacles such as being perceived as unable to speak English, sales people questioning her ability to afford the goods in the store, or simply enduring negative looks or glances from people who are not accepting of another language being spoken in the community. If the speaker chooses to use English, the majority language, she is less likely to encounter these same types of negative social ramifications. Choosing to use Spanish, in this context, can lead to increased stress.

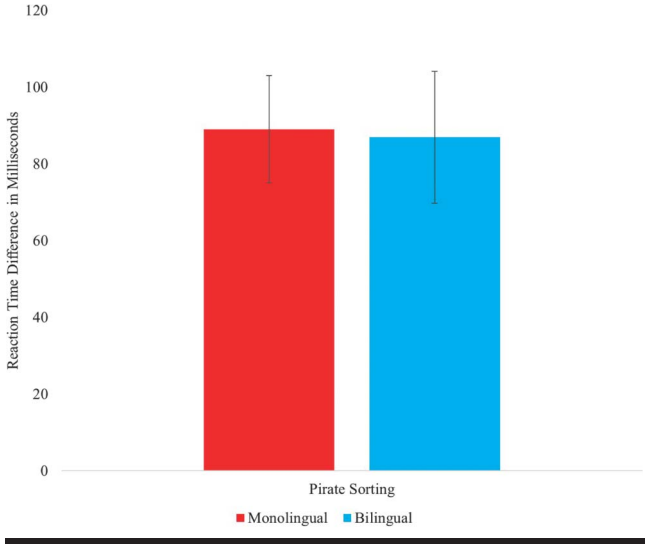
However, this stress does not only manifest in social environments. In Arizona, for example, not only are there policies set in place against teaching in Spanish, but there are also policies that have been discriminatory against individuals from Hispanic or Latino/a backgrounds. These policies affect children. Perceptions among fifth-grade Hispanic students in Arizona were measured and 59% reported

Figure 8. Means and standard errors on reliable tasks measuring accuracy. Bars represent standard errors. *Significant.



that they perceived some discrimination against them (Kulis et al., 2009). This finding is key, considering that studies have shown that daily experiences of perceived discrimination predict psychological distress, major depression, and generalized anxiety (DeGarmo & Martinez, 2006). Importantly, a meta-analysis examining the effect of racism on mental health found that racism was associated with increased psychological stress that was not mediated by age (Paradies et al., 2015). Acculturation stress is the stress that is associated with the expectation that one must fit into the majority culture and 47% of the fifth graders in Kulis et al. (2009) experienced acculturation stress. Acculturation stress directly affects language choice and usage.

Figure 9. Means and standard errors on tasks measuring reaction time differences. Bars represent standard error.



Many of the studies that have shown a bilingual advantage come from places like Canada or Europe, where bilingualism or even multilingualism is not only supported, but expected. Even in these cultures, there can be examples of language use potentially leading to stress. There are two examples from Spain that might show support for the stress hypothesis. Antón et al. (2014) and Duñabeitia et al.'s (2014) studies, which had samples of 360 and 540, respectively, and used multiple measures of the domains they examined, did not find a bilingual advantage. Spain is a culture where there are stark differences in the political and sociolinguistic perceptions of Spanish and Basque, the two languages used in these studies. For decades, it has been documented that there are inherent sociolinguistic differences in these languages. In 1987, Ros, Cano, and Huici noted that Castilian Spanish had the highest level of status, demographics, and institutional support, whereas Basque was considered of medium status, with low demographics,

Table 7. Results of the Bayesian independent-samples *t* test on measures of executive function.

Executive function domain	Task	*Bayes factor in favor of alternative hypothesis*[a]	*Evidence for or against the bilingual advantage hypothesis*[b]
Switching	Pirate Sorting	0.19	AGAINST (substantial evidence for null)
Updating	Number Updating	0.20	AGAINST (substantial evidence for null)
	N-Back Auditory	2694.27	AGAINST (decisive evidence for monolingual advantage)
	N-Back Visual	3.72	AGAINST (substantial evidence for monolingual advantage)

Note. The table does not include the inhibition tasks as they had unacceptable reliability. Please see Table 6.

[a]The Bayes factor for the Bayesian independent-samples *t* test specifies the ratio of the results under the alternative hypothesis versus the null hypothesis. [b]The qualitative labels for the Bayes factor results are based on those by Jeffreys (1961) as modified by Wetzels et al. (2011).

and medium institutional support. It may be the case that these linguistic differences that are tied to cultural and political beliefs may add an additional layer of stress for bilinguals in Spain, as using each of their languages carries with it more than purely linguistic differences.

Task Reliability

Another potential factor that may contribute to the differences found among studies may be task reliability. We reported the reliability of the tasks that we used and chose not to use the tasks that did not meet the reliability criterion. As predicted by the work of Jensen (1965) and others, there was lower reliability on measures of difference (e.g., congruent vs. incongruent trial reaction time performance), like the Stroop task. This was most pronounced for our tasks measuring inhibition. Although we chose not to report the findings from our inhibition measures, many researchers use these tasks without ever reporting (or checking) the reliability (Green et al., 2016). Many may assume that, because these tasks are so frequently used, they are acceptable. Poor reliability could easily lead to variability in the findings across studies. It was difficult to determine if the low reliability we found was particular to our version of these tasks or is more prevalent across this literature. It is important to think about these findings and the reliability of tasks, considering that in 2017 alone, over 9,000 publications used a form of the Stroop task, which is a classic task in the psychology and cognitive science fields, to measure behavior. Only one of the studies we reviewed (Engel de Abreu, 2011) reported reliability statistics for their tasks, and we did not have overlapping tasks.

As clinicians and researchers, we are mindful of selecting tests with strong psychometric properties (e.g., Plante & Vance, 1994) and use data-driven cut scores (e.g., Spaulding, Plante, & Farinella, 2006) to ensure accurate diagnoses. In evaluation, we often strive to use measures with strong sensitivity and specificity to make sure that the groups that we are evaluating can be accurately differentiated from one another. This same principle of using tasks with strong psychometric properties should also apply to the tasks we use to answer research questions. If the measures we use are unreliable, so are the conclusions we draw from them. We need to consider this when we are designing, modifying, and using tasks to test empirical research questions.

Table 8. Correlation between reliable executive function tasks, with significant correlations ($p < .05$) marked with an asterisk.

Task	1	2	3	4
1. Switching: Pirate Sorting	—			
2. Updating: Number Updating	–.01	—		
3. Updating: *N*-Back Auditory	–.10	.16*	—	
4. Updating: *N*-Back Visual	–.03	.27*	.35*	—

Conclusions

Our results more strongly support the null hypothesis—that there are no between-groups differences—than the alternative hypothesis of a bilingual advantage in executive functions. We need to be specific about the people to whom this outcome applies. In our case, the comparison was between typically developing monolingual English and bilingual Spanish–English second graders from a subtractive bilingual environment. We feel most confident about this conclusion for the domain of updating, which had multiple, correlated measures with good reliability. This suggests that the tasks we used were reliably testing the same construct. We have confidence in our measure of shifting, but ideally would like to have another reliable measure of shifting to get a more comprehensive assessment of this domain. We are least confident about the inhibition domain due to the lower reliability of these correlated measures. In conclusion, our findings do not rule out the possibility that a bilingual advantage exists in some children. However, they were not evident in our sample. Based on our findings, differences in executive function would not be a key factor to consider between monolingual and bilingual children whom we serve. However, our findings do lead us to reconsider other effects within each population that may be leading to learning differences and cognitive capabilities. Ideally, future research will use multiple, reliable measures to examine multiple domains of executive function in a well-described population of bilingual children. Hopefully, future work will also be able to more systematically examine the influence of the cultural context of language use and how it may affect executive function development.

Acknowledgments

This work was funded by Grant #R01 DC010784 (with Shelley Gray, PI) from the National Institute on Deafness and Other Communication Disorders, and the first author was supported by diversity supplement 3R01DC010784-04S1, also from the National Institute on Deafness and Other Communication Disorders. We are deeply grateful to the staff, research associates, school administrators, teachers, children, and families who participated. Muchas gracias a las familias que nos ayudaron a completar esta investigación para aprender mas sobre lo que significa ser bilingüe. Apreciamos su tiempo. Key personnel included (in alphabetical order) Shara Brinkley, Gary Carstensen, Cecilia Figueroa, Karen Guilmette, Trudy Kuo, Bjorg LeSueur, Annelise Pesch, and Jean Zimmer. Many students also contributed to this work including (in alphabetical order) Lauren Baron, Alexander Brown, Nora Schlesinger, Nisha Talanki, and Hui-Chun Yang.

References

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80,* 207–245.

Alt, M., Arizmendi, G. D., & DiLallo, J. N. (2016). The role of socioeconomic status in the narrative story retells of school-aged English language learners. *Language, Speech, and Hearing Services in Schools, 47,* 313–323.

Alt, M., Hogan, T., Green, S., Gray, S., Cabbage, K., & Cowan, N. (2017). Word learning deficits in children with dyslexia. *Journal of Speech, Language, and Hearing Research, 60,* 1012–1028.

Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., … Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology, 5,* 1–12.

Antoniou, K., Grohmann, K. K., Kambanaros, M., & Katsos, N. (2016). The effect of childhood bilectalism and multilingualism on executive control. *Cognition, 149,* 18–30.

Arnsten, A. (1998). The biology of being frazzled. *Science, 280,* 1711–1712.

Baddeley, A. (1998). The central executive: A concept and some misconceptions. *Journal of the International Neuropsychological Society, 5,* 523–526.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8,* 47–89.

Barac, R., Moreno, S., & Bialystok, E. (2016). Behavioral and electrophysiological differences in executive control between monolingual and bilingual children. *Child Development, 87,* 1277–1290.

Barragan, B., Castilla-Earls, A. P., Martinez-Nieto, L., Restrepo, M. A., & Gray, S. (2018). Performance of low-income dual language learners attending English-only schools on the Clinical Evaluation of Language Fundamentals–Fourth Edition, Spanish. *Language, Speech, and Hearing Services in Schools,* 292–305. https://doi.org/10.1044/2017_LSHSS-17-0013

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

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19,* 290–303.

Bialystok, E., Kroll, J. F., Green, D. W., MacWhinney, B., & Craik, F. I. (2015). Publication bias and the validity of evidence: What's the connection? *Psychological Science, 26,* 944–946.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science, 7,* 325–339.

Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition, 112,* 494–500.

Blair, C., Granger, D., & Peters Razza, R. (2005). Cortisol reactivity is positively related to executive function in preschool children attending Head Start. *Child Development, 76,* 554–567.

Blom, E., Boerma, T., Bosma, E., Cornips, L., & Everaert, E. (2017). Cognitive advantages of bilingual children in different sociolinguistic contexts. *Frontiers in Psychology, 8,* 1–12.

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer [Computer program]. Version 5.3.64. Retrieved from http://www.praat.org/

Bonifacci, P., Giombini, L., Bellocchi, S., & Contento, S. (2011). Speed of processing, anticipation, inhibition and working memory in bilinguals. *Developmental Science, 14,* 256–269.

Brownell, R. (2001). *Expressive One-Word Picture Vocabulary Test–Spanish-Bilingual Edition*. Novato, CA: Academic Therapy Publications.

Cabbage, K., Brinkley, S., Gray, S., Alt, M., Cowan, N., Green, S., … Hogan, T. (2017). Assessing Working Memory in Children: The Comprehensive Assessment Battery for Children–Working Memory (CABC-WM). *Journal of Visual Experiments, 124,* e55121.

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11,* 282–298.

Chen, S. H., Zhou, Q., Uchikoshi, Y., & Bunge, S. A. (2014). Variations on the bilingual advantage? Links of Chinese and English proficiency to Chinese American children's self-regulation. *Frontiers in Psychology, 5,* 1069.

Cowan, N., Hogan, T. P., Alt, M., Green, S., Cabbage, K. L., Brinkley, S., & Gray, S. (2017). Short-term memory in childhood dyslexia: Deficient serial order in multiple modalities. *Dyslexia, 23,* 209–233.

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science, 26,* 99–107.

DeGarmo, D., & Martinez, C. (2006). A culturally informed model of academic well-being for Latino youth: The importance of discriminatory experiences and social support. *Family Relations, 55,* 267–278.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4–12 years old. *Science, 333,* 959–964.

Diaz, R. M. (1985). Bilingual cognitive development: Addressing three gaps in current research. *Child Development, 56,* 1376–1388.

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: Myth or reality. *Experimental Psychology, 61,* 234–251.

DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Reid, R., Ikeda, M. J., & McGoey, K. E. (1998). Parent ratings of attention deficit/hyperactivity disorder symptoms: Factor structure and normative data. *Journal of Psychopathology and Behavioral Assessment, 20,* 83–102.

Engel de Abreu, P. M. J. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory, 19,* 529–537.

Engel de Abreu, P. M., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism enriches the poor: Enhanced

cognitive control in low-income minority children. *Psychological Science, 23,* 1364–1371.

Filippi, R., Morris, J., Richardson, F. M., Bright, P., Thomas, M. S., Karmiloff-Smith, A., & Marian, V. (2015). Bilingual children show an advantage in controlling verbal interference during spoken language comprehension. *Bilingualism: Language and Cognition, 18,* 490–501.

Friedman, N. P. (2016). Research on individual differences in executive functions. *Linguistic Approaches to Bilingualism, 6,* 535–548.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2)*. Circle Pines, MN: AGS.

Gray, S., Green, S., Alt, M., Hogan, T., Kuo, T., Brinkley, S., & Cowan, N. (2017). The structure of working memory in young children and its relation to intelligence. *Journal of Memory and Language, 92,* 183–201.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1,* 67–81.

Green, S. B., Yang, Y., Alt, M., Brinkley, S., Gray, S., Hogan, T., & Cowan, N. (2016). Use of internal consistency coefficients for estimating reliability of experimental task scores. *Psychonomic Bulletin & Review, 23,* 750–763.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review, 18,* 625–658.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Jensen, A. R. (1965). Scoring the Stroop test. *Acta Psychologica, 24,* 398–408.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review, 17,* 213–233.

Kapa, L. L., & Colombo, J. (2013). Attentional control in early and later bilingual children. *Cognitive Development, 28,* 233–246.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children–Second Edition (K-ABC2)*. Circle Pines, MN: AGS.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin, 99*(3), 422–431.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55,* 352–358.

Kulis, S., Marsiglia, F. F., & Nieri, T. (2009). Perceived ethnic discrimination versus acculturation stress: Influences on substance use among Latino youth in the southwest. *Journal of Health and Social Behavior, 50,* 443–459.

Lambert, W. E. (1973, November). *Culture and language as factors in learning and education*. Paper presented at the Annual Learning Symposium on Cultural Factors in Learning, Bellingham, WA.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*(2), 202–220.

Mahadi, T. S. T., & Jafari, S. M. (2012). Language and culture. *International Journal of Humanities and Social Science, 2,* 230–235.

Martinez, C. (2006). Effects of differential family acculturation on Latino adolescent substance use. *Family Relations, 55,* 306–317.

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,* 49–100.

Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology, 114,* 187–202.

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science, 10,* 719–726.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9,* 353–383.

Norman, D. A., & Shallice, T. (1986). Attention to action. Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research* (Vol. IV). New York, NY: Plenum Press.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 411–421.

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.

Paradies, Y., Ben, J., Denson, N., Elias, A., Priest, N., Pieterse, A., . . . Gee, G. (2015). Racism as a determinant of health: A systematic review and meta-analysis. *PloS One, 10,* 1–48.

Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied, 76,* 1–23.

Pechtel, P., & Pizzagalli, D. A. (2011). Effects of early life stress on cognitive and affective function: An integrated review of human literature. *Psychopharmacology, 214,* 55–70.

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25,* 15–24.

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology, 113,* 535–551.

Ros, M., Cano, J. I., & Huici, C. (1987). Language and intergroup perception in Spain. *Journal of Language and Social Psychology, 6,* 243–259.

Ross, J., & Melinger, A. (2017). Bilingual advantage, bidialectal advantage or neither? Comparing performance across three tests of executive function in middle childhood. *Developmental Science, 20*(4), e12405.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4)*. Circle Pines, MN: AGS.

Semel, E., Wiig, E. H., & Secord, W. A. (2006). *Clinical Evaluation of Language Fundamentals–Fourth Edition, Spanish (CELF-4 Spanish)*. San Antonio, TX: Pearson.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 298,* 199–209.

Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools, 37,* 61–72.

Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Telles, E. E., & Murguia, E. (1990). Phenotypic discrimination and income differences among Mexican Americans. *Social Science Quarterly, 71,* 682.

Torgesen, J., Wagner, R., & Rashotte, C. (2012). *Test of Word Reading Efficiency–Second Edition (TOWRE-2)*. Circle Pines, MN: AGS.

U.S. Census Bureau. (2015). *Language spoken at home, 2011–2015 American Community Survey 5-Year Estimates*. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_15_5YR_S1601&prodType=table

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspectives on Psychological Science, 6*(3), 291–298.

Williams, K. T. (2007). *Expressive Vocabulary Test–Second Edition (EVT-2)*. London, UK: Pearson Assessments.

Woodcock, R. W. (2011). *Woodcock Reading Mastery Tests–Third Edition*. Bloomington, MN: Pearson.

Yang, S., & Yang, H. (2016). Bilingual effects on deployment of the attention system in linguistically and culturally homogeneous children and adults. *Journal of Experimental Child Psychology, 146,* 121–136.

Appendix (p. 1 of 2)

Additional Analyses Using a Covariate or Matching

The results below are those of the full data set for all reliable measures with maternal level of education (MLE), a proxy for socioeconomic status, used as a covariate.

Executive function domain	Task	Adjusted monolingual *M* (*SEM*)	Adjusted bilingual *M* (*SEM*)	*F*	*p*	$\eta^{2partial}$	*Was MLE significant?*
Shifting	Pirate Sorting	89.57 (14.45)	79.27 (23.05)	0.12	0.720	< .001	NO
Updating	Number Updating	85.57 (2.22)	87.04 (3.50)	0.10	0.740	< .001	NO
	N-Back Auditory*	82.73 (1.31)	75.07 (2.05)	8.52	0.003	0.03	NO
	N-Back Visual	76.14 (1.62)	70.01 (2.56)	3.56	0.060	0.01	NO

Note. SEM = standard error of the mean.

*Statistically significant between-groups difference using Bonferroni corrections for multiple comparisons.

The results below are those of the full data set for reliable measures with the Kaufman Assessment Battery for Children–Second Edition (K-ABC2) score, a measure of nonverbal intelligence, used as a covariate.

Executive function domain	Task	Adjusted monolingual *M* (*SEM*)	Adjusted bilingual *M* (*SEM*)	*F*	*p*	$\eta^{2partial}$	*Was K-ABC2 significant?*
Shifting	Pirate Sorting	96.48 (13.49)	71.47 (19.87)	1.02	0.31	0.004	YES
Updating	Number Updating	84.13 (2.08)	89.92 (3.06)	2.30	0.12	0.009	YES
	N-Back Auditory*	82.84 (1.31)	74.18 (1.91)	13.21	< .001	0.050	NO
	N-Back Visual	74.68 (1.49)	72.57 (2.21)	0.59	0.44	0.002	YES

Note. SEM = standard error of the mean.

*Statistically significant between-groups difference using Bonferroni corrections for multiple comparisons.

Appendix (p. 2 of 2)

Additional Analyses Using a Covariate or Matching

The results below represent 57 pairs of students individually matched on maternal level of education (monolingual *X* = 13.96, *SD* = 1.60; bilingual *X* = 13.50, *SD* = 2.01), sex, and age (monolingual *X* = 7;8, *SD* = 0;5; bilingual *X* = 7;8, *SD* = 0;5). Results below are for paired *t* tests. We used paired *t* tests to deal with the dependency that comes from individual matching (see Kenny & Judd, 1986).

Central executive category	Task	Monolingual *M* (*SD*)	Bilingual *M* (*SD*)	*t*[a]	*p*[b]	*d*[a]	*Bayes factor in favor of alternative hypothesis*[c]	*Evidence for or against the bilingual advantage hypothesis*[d]
Shifting	Pirate Sorting	93.33 (198.79)	80.51 (137.47)	0.36	0.72	0.05	0.20	AGAINST (substantial evidence for null)
Updating	Number Updating	85.54 (28.45)	88.64 (22.15)	–0.58	0.56	–0.08	0.22	AGAINST (substantial evidence for null)
	N-Back Auditory	81.57 (14.38)	73.04 (19.86)	2.41	0.01	0.33	2.83	AGAINST (anecdotal evidence for monolingual advantage)
	N-Back Visual	73.99 (18.34)	71.27 (22.98)	0.69	0.49	0.09	0.23	AGAINST (substantial evidence for null)

Note. The table does not include the inhibition tasks as they had unacceptable reliability. Please see Table 6.

[a]For the paired-samples *t* test and the *d* effect size statistic, a negative value indicates that the bilingual group had the higher mean on a measure of executive function. [b]The *p* values for the paired-samples *t* test should be compared to an α of .05/7 = .007 following the Bonferroni method. [c]The Bayes factor for the Bayesian paired-samples *t* test specifies the ratio of the results under the alternative hypothesis versus the null hypothesis. [d]The qualitative labels for the Bayes factor results are based on those by Jeffreys (1961) as modified by Wetzels et al. (2011).

The results below represent 72 pairs of students individually matched on nonverbal intelligence using K-ABC scores (monolingual *X* =108.88, *SD* = 11.64; bilingual *X* = 107.86, *SD* =11.65), sex, and age (monolingual *X* =7;8, *SD* = 0;4; bilingual *X* = 7;8, *SD* =0;4). Results below are for paired *t* tests. We used paired *t* tests to deal with the dependency that comes in from individual matching (see Kenny & Judd, 1986).

Central executive category	Task	Monolingual *M* (*SD*)	Bilingual *M* (*SD*)	*t*[a]	*p*[b]	*d*[a]	*Bayes factor in favor of alternative hypothesis*[c]	*Evidence for or against the bilingual advantage hypothesis*[d]
Shifting	Pirate Sorting	107.76 (198.93) (178.50)	88.03 (157.36)	0.65	0.51	0.08	0.20	AGAINST (substantial evidence for null)
Updating	Number Updating	80.94 (32.72)	86.76 (23.27)	–1.15	0.25	–0.14	0.31	AGAINST (substantial evidence for null)
	N-Back Auditory	84.36 (13.38)	71.28 (20.29)	4.06	< .001	0.50	234.30	AGAINST (decisive evidence for monolingual advantage)
	N-Back Visual	74.12 (18.86)	70.86 (21.87)	0.92	0.35	0.16	0.25	AGAINST (substantial evidence for null)

Note. The table does not include the inhibition tasks as they had unacceptable reliability. Please see Table 6.

[a]For the paired-samples *t* test and the *d* effect size statistic, a negative value indicates that the bilingual group had the higher mean on a measure of executive function. [b]The *p* values for the paired-samples *t* test should be compared to an α of .05/7 = .007 following the Bonferroni method. [c]The Bayes factor for the Bayesian paired-samples *t* test specifies the ratio of the results under the alternative hypothesis versus the null hypothesis. [d]The qualitative labels for the Bayes factor results are based on those by Jeffreys (1961) as modified by Wetzels et al. (2011).