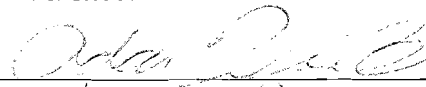


RELATIONS BETWEEN EXECUTIVE FUNCTIONING, SECOND LANGUAGE  
FLUENCY, AND EXTERNALIZING BEHAVIOR PROBLEMS IN EARLY  
CHILDHOOD

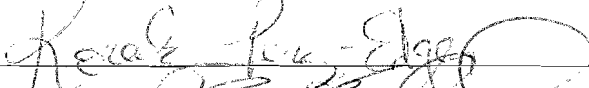
by

Lindsey A. Hutchison  
A Dissertation  
Submitted to the  
Graduate Faculty  
of  
George Mason University  
in Partial Fulfillment of  
The Requirements for the Degree  
of  
Doctor of Philosophy  
Psychology

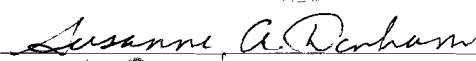
Committee:

  
\_\_\_\_\_

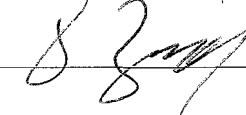
Director

  
\_\_\_\_\_

Department Chairperson

  
\_\_\_\_\_

Program Director

  
\_\_\_\_\_

Dean, College of Humanities  
and Social Sciences

Date:

  
\_\_\_\_\_

Spring Semester 2012  
George Mason University  
Fairfax, VA

Relations between Executive Functioning, Second Language Fluency, and Externalizing  
Behavior Problems in Early Childhood

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of  
Philosophy at George Mason University

by

Lindsey A. Hutchison  
Master of Arts  
George Mason University, 2009

Director: Adam Winsler, Professor  
Department of Psychology

Spring Semester 2012  
George Mason University  
Fairfax, VA

## DEDICATION

This dissertation is dedicated to my parents, John Michael Hutchison and Elaine Carol Hutchison. Without their continued love and support, this project would not have been possible.

## ACKNOWLEDGEMENTS

I would like to thank the many individuals that made this project possible. My team of graduate and undergraduate research assistants helped to collect, code, and enter all data. My parents provided invaluable emotional support. The members of my committee and my psychology lab provided helpful feedback. Finally, tremendous thanks go out to my advisor, Adam Winsler, for his unwavering support of this project.

## TABLE OF CONTENTS

	Page
List of Tables .....	vii
List of Figures.....	viii
List of Abbreviations .....	ix
Abstract.....	x
INTRODUCTION.....	1
Why Study Bilingualism and Externalizing Behavior Problems? .....	3
Defining Executive Function .....	5
The Development of Executive Functioning.....	9
Inhibitory Control .....	9
Cognitive Flexibility or Set Shifting .....	13
Planning Skills .....	14
Executive Functioning for Children with Externalizing Behavior Problems.....	16
Executive Functioning for Children Who Speak More Than One Language .....	21
Bilingualism, Language Development, and Externalizing Behavior Problems .....	33
Rationale for the Current Study.....	38
Method .....	44
Participants.....	44
Procedure .....	48
Training of Research Assistants .....	55
Measures .....	56
Receptive Language Skills.....	56
Expressive Language Skills .....	58
Transcribing and Coding of Elicited Narratives.....	60
Language Experience and Exposure .....	61
Determining Child Language Group.....	64
Calculating Degree of Bilingualism .....	66

Executive Functioning: Cognitive Flexibility.....	67
Executive Functioning: Interference Control.....	70
Executive Functioning: Inhibition of a Prepotent Response.....	73
Executive Functioning: Planning Skills .....	75
Executive Functioning: Inhibitory Control – Replacement of a Dominant Response with a Less-Dominant Response .....	78
Executive Functioning: Parent Report .....	81
Externalizing Behavior Problems .....	82
Results.....	84
Exploratory Data Analyses .....	84
Covariate Identification.....	86
Language Group Differences in Demographic Variables.....	89
Data Reduction .....	90
Research Question 1 .....	91
Group Differences in Inhibitory Control.....	92
Group Differences in Cognitive Flexibility .....	94
Group Differences in Planning/Problem-Solving and Parent Report of EF .....	95
Research Question 2 .....	96
Research Question 3 .....	97
Research Question 4.....	98
Research Question 5.....	98
Research Question 6.....	100
EF as Mediator of Language Group – Behavior Problems Relationship.....	100
Language Group as Moderator of EF – Behavior Problems Relationship .....	103
Degree of Bilingualism as Moderator of EF – Behavior Problems Relationship ....	113
Discussion.....	118
General Findings on EF Development and Theory.....	118
Findings on Second Language Fluency and EF .....	122
Findings on Second Language Fluency and Behavior .....	131
Findings on EF and Behavior .....	132
Findings on the Interplay of Language, EF, and Behavior .....	133
Strengths and Contributions of the Current Study .....	136
Limitations and Considerations for Future Research.....	139

General Conclusions .....	145
References.....	150

## LIST OF TABLES

Table	Page
Table 1. Descriptive information, overall and by language group.....	46
Table 2. Description of all measures of EF, behavior, and language. ....	51
Table 3. Bivariate correlations between all EF outcome measures. ....	88
Table 4. Means and SDs for EF outcome measures by language group. ....	92
Table 5. Correlations between externalizing behavior problems (CBCL) and EF measures, with and without outliers included. ....	99
Table 6. Bivariate correlations between externalizing behavior problems score and all EF outcome measures, separately by language group. ....	105
Table 7. Moderated regression results for language group as a moderator of the relationship between EF (two types of inhibitory control) and behavior problems.....	107
Table 8. Moderated regression results for language group as a moderator of the relationship between parent report of EF (BRIEF MI) and behavior problems.....	111
Table 9. Bivariate correlations between externalizing behavior problems score and all EF measures, separately for those “high” and “low” on degree of bilingualism. ....	114
Table 10. Moderated regression results for degree of bilingualism as a moderator of the relationship between inhibition of a prepotent response (GNG) and behavior problems (CBCL). ....	115



## LIST OF FIGURES

Figure	Page
Figure 1. Order of experimental procedures for parent and child.....	50
Figure 2. Language group as a moderator of the relationship between inhibitory control (HTKS score) and externalizing behavior problems (CBCL).....	108
Figure 3. Language group as a moderator of the relationship between interference control and externalizing behavior problems. ....	109
Figure 4. Language group as a moderator of the relationship between metacognitive functioning (BRIEF) and externalizing behavior problems (CBCL). ....	112
Figure 5. Degree of bilingualism as a moderator of the relationship between inhibition of a prepotent response (GNG commission errors) and externalizing behavior problems (CBCL).....	116
Figure 6. Degree of bilingualism as a moderator of the relationship between metacognitive skills (BRIEF) and externalizing behavior problems (CBCL). ....	117

## LIST OF ABBREVIATIONS

Attention Deficit Hyperactivity Disorder .....	ADHD
Behavior Rating Inventory of Executive Function .....	BRIEF
Child Behavior Checklist.....	CBCL
Dimensional Change Card Sort.....	DCCS
Dual Language Learner.....	DLL
Executive Functioning .....	EF
Go/No-Go task .....	GNG
Head-Toes-Knees-Shoulders task .....	HTKS
Peabody Picture Vocabulary Test .....	PPVT
Test de Vocabulario en Imágenes Peabody .....	TVIP
Tower of London.....	TOL
United States.....	U.S.

## ABSTRACT

### RELATIONS BETWEEN EXECUTIVE FUNCTIONING, SECOND LANGUAGE FLUENCY, AND EXTERNALIZING BEHAVIOR PROBLEMS IN EARLY CHILDHOOD

Lindsey A. Hutchison, Ph.D.

George Mason University, 2012

Dissertation Director: Dr. Adam Winsler

The current study explores relations between executive functioning (EF), degree of bilingualism, and externalizing behavior problems in a sample ( $N = 79$ ) of 5- to 7-year-old monolingual ( $n = 33$ ) and bilingual ( $n = 46$ ) children. The bilingual group included both children who were fully fluent in two languages (balanced bilinguals;  $n = 17$ ) and children who were still learning their second language (Dual Language Learners; DLLs;  $n = 29$ ). The main components of EF included inhibitory control, cognitive flexibility, and planning/problem-solving skills. Parents and children came into the lab for a one-time, two-hour session. Parents completed surveys on children's language background, EF, and behavior problems while children worked with the researcher in a separate room. EF was measured with a go/no-go task (GNG), the Head-Toes-Knees-Shoulders task (HTKS), the Dimensional Change Card Sort (DCCS), the Simon task, and the Tower of London (TOL). Children's language proficiency in English and Spanish was measured

with direct assessments. Results revealed no language group differences for simple inhibitory control (i.e. tasks that were more response-based in nature; GNG, HTKS), but the “bilingual advantage” was demonstrated for more complex inhibitory control (i.e. task that involved control of attention; Simon) and cognitive flexibility (DCCS).

Bilinguals also demonstrated superior planning/problem-solving skills (TOL), an area with little prior research. Contrary to prior research, there was some evidence that DLLs had an advantage over monolinguals in interference control and cognitive flexibility.

There was no evidence that EF mediated the relationship between language group and behavior problems. There was some evidence that the relationships between EF and

behavior problems were stronger for balanced bilinguals compared to DLLs and

monolinguals. Findings have important implications in light of the growing population of Spanish-English balanced bilinguals and DLLs in the U.S.

## INTRODUCTION

In today's high-stakes academic environment, more demands are being placed on students at younger ages to focus attention for sustained periods, shift flexibly from one task to another, plan and organize behavior, and set and follow through with long-term goals (Meltzer, 2010). Such demands require children to engage their developing executive functioning (EF) abilities in order to succeed (Meltzer, 2010). The term "executive functioning" or "executive functions" (EFs) is a broad construct, encompassing several different components or abilities (Anderson, Anderson, Jacobs, & Smith, 2008; Garon, Bryson, & Smith, 2008; Pennington, 1994; Pennington & Ozonoff, 1996; Stuss, 1992; Welsh, Pennington, & Groisser, 1991). As a construct, EF is difficult to capture, and therefore suffers from poor and inconsistent definition (Burgess, 1997; Rabbitt, 1997). However, most researchers agree that EF involves various processes for planning, initiating, and following through with goal-directed behavior (Anderson et al., 2008; Burgess, 1997; Garon et al., 2008; Welsh et al., 1991); it is thought to "free behavior" from the moment-to-moment environment (Brocki & Bohlin, 2004). Theoretical conceptions of EF are explained in more detail below. In the classroom setting, EF involves stopping and thinking before acting, as well as organizing and planning one's behavior (Kaufman, 2010).

EF has utility as a school readiness construct. The various components develop with age (discussed in further detail below), such that preschool children are acquiring many of the higher-order abilities necessary to function in an attentive and engaging manner in kindergarten (Blair, 2002; Luciana & Nelson, 1998). A survey of kindergarten teachers that explored what characteristics teachers view as most important for starting school indicated that 84% thought children should be able to communicate their needs verbally, 76% thought children should be enthusiastic about learning, 60% indicated the abilities to follow directions and not disrupt class were important, 10% thought children should know several letters, and only 7% thought children should be able to count to 20 (Lewit & Baker, 1995). Findings such as these reveal that kindergarten teachers are very concerned with children's ability to communicate, pay attention, and engage in the classroom, as opposed to coming into kindergarten simply with knowledge of numbers and letters. This implies a general concern by early elementary school teachers with children's ability to regulate their own behavior (Blair, 2002), and executive functioning skills are at the crux of self-regulation (Blair, 2002; Kaufman, 2010). Therefore, educators and researchers alike should be concerned with specific factors that influence the development of EF, as well as individual children's EF strengths and weaknesses. In particular, children who speak two languages fluently demonstrate advantages in EF (Bialystok, 2001), while children with externalizing behavior problems demonstrate difficulties with EF (Barkley, 1997). Such children are important to study in light of the growing proportion of children in public schools that are English Language Learners (ELLs; Gandara & Rumberger, 2009), as well as the increasing number of children being

diagnosed with some kind of externalizing behavioral disorder, such as attention-deficit hyperactivity disorder (ADHD; Mandell, Thompson, Weintraub, DeStefano, & Blank, 2005; Visser, Bitsko, Danielson, Perou, & Blumberg, 2010).

### **Why Study Bilingualism and Externalizing Behavior Problems?**

The number of individuals in the United States who do not speak English as a first language has increased drastically in the past several decades. In 2000, the U.S. Census Bureau estimated that individuals who do not speak English as their primary language comprised 18% of the total population – approximately 47 million people (Shin & Bruno, 2003). When these individuals are children in U.S. public schools, they are often referred to as English Language Learners (ELLs) or Dual Language Learners (DLLs), though policy-makers use these labels interchangeably with several other possibilities (i.e. Limited English Proficient; LEP).

With the new wave of immigration in recent years, the percentage of students who are DLLs has increased at a much faster rate than the general student population. For instance, the general population of students increased by about 20% from 1979 to 2006, but those speaking a language other than English increased by 185% in this same time frame (Gandara & Rumberger, 2009). Based on this explosive growth, it is predicted that White students will be the minority in just a few more decades (Garcia & Cuellar, 2006). In general, research on DLLs has indicated that they tend to struggle in U.S. public schools, performing lower than their native English-speaking peers, especially on standardized assessments of math and reading (Moss & Puma, 1995). However, a sizeable body of research has also indicated that children who are fully fluent in two

languages display advantages in EF, including selective attention, set shifting, and inhibition in the presence of conflicting information (Bialystok, 2001; Martin-Rhee & Bialystok, 2008). Less is known, however, about potential EF advantages for children who are not fully fluent in their second language.

Similarly, the number of children diagnosed with disorders characterized by externalizing behavior problems, such as ADHD, has apparently been on the rise (Mandell et al., 2005; Visser et al., 2010). Mandell and colleagues examined a sample of children that had been hospitalized for various disorders between the years 1989 and 2000. They found that the rate of children diagnosed with ADHD increased almost four-fold during this time frame. A more recent study conducted by the Center for Disease Control compared parent-reported rates of ADHD in 2003 to those in 2007 (Visser et al., 2010). During this four-year time frame, the percent of children with an ADHD diagnosis increased from 7.8% to 9.5%, a change of 21.8%. Interestingly, the increase was most pronounced for Hispanic children and children who were non-native English speakers, though the overall numbers were lower for these groups compared to other ethnic and language groups (Visser et al., 2010).

It is important to take more notice of children with behavior problems, as children with poor regulation early in life are more likely to have problems later. Indeed, Winsler, Diaz, Atencio, McCarthy, and Chabay (2000) demonstrated that children who were “hard to manage” at 3.5 years old (identified by teacher ratings of child behavior) were more likely to have attention and behavior problems (as reported by their parents) at 5.5 years old. A large body of research has demonstrated that children who have externalizing



problems and poor self-regulation skills in early childhood are likely to have persistent difficulties later in childhood and/or adolescence, compared to children with better self-regulation and/or infrequent behavior problems early on (Bilancia & Rescorla, 2010; Campbell, 1994; Campbell, 2002; Campbell, Shaw, & Gilliom, 2000; Lavigne et al., 1998; Moreland & Dumas, 2008; Winsler et al., 2000).

Much research has already examined bilingualism and EF, as well as behavior problems and EF, but no studies to date have examined how these three constructs might relate in various ways. As noted above, knowledge of more than one language is related to EF advantages and behavior problems are related to EF deficits. Therefore, in addition to exploration of bivariate relationships between bilingualism and EF, and between EF and behavior, the current paper also examined how EF and behavior problems might relate differently for bilingual vs. monolingual children, and/or whether EF might mediate the relationship between bilingualism and behavior problems. Additionally, very few studies have examined samples of children who are not fully fluent in their second language (DLLs), another strength of the current study.

### **Defining Executive Function**

At the most general level, EF is conceived of as the engagement of conscious, as opposed to automatic, cognitive processes (Burgess, 1997; Rabbitt, 1997). For example, thinking through the steps necessary to solve a new problem (i.e. planning) before actually attempting the problem is a conscious cognitive process, while well-learned behaviors such as reading, walking, and driving are more automatic processes. EF helps individuals to navigate and adapt to novel situations via key functions such as planning,

problem-solving, task initiation, monitoring of one's progress, inhibition of inappropriate or unnecessary responses, working memory, and sustainment and flexibility of attentional resources (Anderson, 2008; Burgess, 1997; Rabbitt, 1997). However, as a construct, EF suffers from inconsistent and unclear theoretical understanding (Miyake et al., 2000). One reason for this poor definition of EF is due to the sheer number of higher-order processes it is thought to encompass, and a lack of consensus on exactly what abilities or components fall under the domain of EF (Brocki & Bohlin, 2004; Pennington & Ozonoff, 1996; Senn, Espy, & Kaufmann, 2004). In their review, for instance, Best, Miller, and Jones (2009) identified over 15 components of EF from prior studies.

Several different theoretical frameworks or models have been proposed to understand and explain EF. Some describe EF as containing a supervisory attention system with many different subsystems or sub-processes (Norman & Shallice, 1986; Shallice & Burgess, 1996; Stuss, Shallice, Alexander, & Picton, 1995), others point out the importance of working memory as central to EF (Baddeley, 1986, 2000), while still others argue for the central importance of behavioral inhibition to EF (Barkley, 1997). Researchers also disagree on whether to conceptualize EF as a unitary construct, or as a set of different but integrated functions. This latter conception of EF as separate, but overlapping, components has received considerable research support in recent years (Anderson, 2002; Best et al., 2009; Miyake et al., 2000).

First, studies have shown that EF components develop at different rates and reach adult-like maturity at different ages (Best et al., 2009). For instance, children become better at controlling impulses (inhibition) throughout the preschool years (Carlson, 2005;

Diamond, 2002; Mischel, Shoda, & Rodriguez, 1989), though planning abilities remain fairly poor through preschool (but with some improvement from 3 to 5 years; Atance & Jackson, 2009), and then start to improve during the elementary school years (Levin et al., 1991; Luciana & Nelson, 1998). The fact that different components have different developmental trajectories provides evidence for some independence of EF components. Second, children with various neuropsychological disabilities show more impairment in certain components of EF over others, also suggesting that the components are at least somewhat independent of one another (Anderson, 2008; Best et al., 2009). Anderson (2002) proposed that the different domains of EF (which he views as information processing, attentional control, cognitive flexibility, and goal setting) work together to act as a larger executive control system.

Miyake and colleagues (2000) have also found support for EF components as separate but integrated. In their study, Miyake et al. (2000) examined inhibition, set shifting (shifting attention flexibly from one set of information or tasks to another), and updating/monitoring of working memory, arguing that these three EFs are the building blocks for more complex EFs, such as planning. Through their confirmatory factor analysis approach, they did indeed find support for these three separate, though correlated, EF components. Miyake et al. concluded that EF is best thought of as separate components that overlap and relate to one another in functionally important ways. The components described by Miyake and colleagues have also been found frequently in other factor analytic studies (Garon et al., 2008; Lehto, Juujarvi, Kooistra, & Pulkkinen, 2003). The current paper will take Anderson (2002) and Miyake et al.'s more integrated

view of EF as one construct with multiple inter-related components. Specifically, I examined inhibition, cognitive flexibility/set-shifting, and planning. Although working memory is typically considered a foundational EF, it will not be included as a separate construct in the current paper as it is inherent in other higher-order EFs (planning, set shifting).

EF is carried out primarily by the prefrontal cortex (PFC), as evidenced by results of brain lesion studies. Many aspects of goal-directed behavior are impaired in individuals with lesions to the PFC (Pennington & Ozonoff, 1996). The development of EF (and frontal brain regions more broadly) is an experience-expectant process (Carlson, 2003; Garon et al., 2008), requiring environmental input to reach full developmental maturation (Couperus & Nelson, 2006; Nelson & Bloom, 1997). Further, the development of the PFC is protracted over several years, continuing well into late adolescence and early adulthood (Gogtay et al., 2004), leaving ample time for environmental impact. Interestingly, there is also evidence that development of the PFC is linked to heritability more closely than other areas of the brain (Kalbfleisch, 2004). One can conceptualize this development as involving transactional processes between the child, his/her biology, and his/her experience (Rutter et al., 1997), where this interplay may result in EF deficits for some and advantages for others. As mentioned previously, experience with two languages is related to EF advantages (Bialystok, 2001), while externalizing behavior problems are related to disadvantages (Brophy, Taylor, & Hughes, 2002; Raaijmakers et al., 2008). I will now present a summary of the typical function and development of my three chosen EFs (inhibitory control, cognitive flexibility, and

planning) during the early school-age years, as well as what patterns have been noted for bilingual children and those with behavior problems. Last, I examine the limited research on bilingualism and behavior problems.

## **The Development of Executive Functioning**

### **Inhibitory Control**

Successful inhibitory control involves the ability to manage responding in a variety of situations. Inhibitory control is represented in the management of outwardly observable motor behaviors, such as when a child delays gratification, resists temptation, and controls impulses (Harnishfeger, 1995). The current study examined three kinds of inhibitory control: 1) inhibition of a prepotent response – a very simple form of inhibition that involves suppressing a dominant response when cued to do so (Nigg, 2000); 2) suppressing a dominant response *and replacing* it with a less common response; and 3) interference control, which involves making an appropriate/correct response by ignoring a distracting feature of a stimulus (e.g. location) and instead attending to a target feature of that stimulus (e.g. color; Nigg, 2000). Each of these types of inhibition is explained in more detail below.

An example of inhibiting a prepotent response is seen with a go/no-go task (Nigg, 2000), where participants are shown different stimuli on a screen, such as a red square and a blue square. Whenever the red square (“go” stimulus) appears, they are to respond as quickly as possible by pushing a button, and when the blue square appears (“no-go” stimulus), they should *not* push the button. The majority of trials are “go trials” in order to build a *prepotent* response, where the individual is expecting to engage in motor

behavior by pushing the button (Brocki & Bohlin, 2004). “No-go trials” are interspersed throughout the task (usually 25% to 30% of trials). Researchers typically look at the percent of commission errors – when a participant pushes the button on a no-go trial – as a measure of inhibitory control (Berlin & Bohlin, 2002; Brocki & Bohlin, 2004).

The second type of inhibitory control – replacing a dominant response with a less common response – is exemplified with tasks like the Day-Night Stroop-like task (Gerstadt, Hong, & Diamond, 1994) and the Head-Toes-Knees-Shoulders (HTKS) task (Ponitz, McClelland, Matthews, & Morrison, 2009). The Day-Night Stroop-like task requires children to say “night” to a picture of the sun and “day” to a picture of the moon (Gerstadt et al., 1994). Children must inhibit the habitual (prepotent) response of saying “day” when they see the sun and instead say the opposite. With the HTKS, the child must do the opposite of what the experimenter says (e.g. touch his/her toes when experimenter says “touch your head” or touch his/her knees when experimenter says “touch your shoulders) by suppressing a dominant/ competing response and performing a different response instead (Ponitz et al., 2009). [The HTKS is discussed further in the Study Rationale and Method sections below.] It is important to note that, for both of these kinds of tasks, the stimulus presented contains only one perceptual or verbal feature (e.g. picture of “sun” or experimenter says “touch your toes”) and the child must choose between competing response options (Martin-Rhee & Bialystok, 2008). In order to be successful, the child must choose the less-dominant (less common) response. However, no interference or conflict is present (as would occur when presented with a stimulus with multiple features). Therefore, this kind of inhibition, as well as inhibition of a prepotent

response, discussed above, are more response-based in nature, and do not appear to require attentional control (i.e., interference control, discussed next) (Martin-Rhee & Bialystok, 2008).

An example of interference control is seen with the classic Stroop task (Stroop 1935), where the participant must say the *ink color* of a list of words, and the words themselves are the names of colors (i.e. the word “green” written in red ink and the correct response is “red”). The participant must suppress the interfering information (the word itself), and instead say out loud the correct piece of information (the ink color). The participant never intends to read out the actual word, but that information is present and distracting. Unlike the first two forms of inhibition that are more response-based in nature, tasks tapping into interference control also require attentional/cognitive control. The stimuli in these tasks have two conflicting features present at the same time on a given trial (such as a color word and the color of the ink) and require a participant to attend to only one feature in order to respond correctly (Bunge, Dudukovic, Thomason, Vaidya, & Gabrieli, 2002; Martin-Rhee & Bialystok, 2008). Therefore, interference control can be conceptualized as a more complex inhibitory control process requiring both cognitive control (controlling attention) *and* a choice between competing responses (e.g. in Stroop task, choice between reading color word or saying ink color).

Inhibitory control is developing from infancy (Diamond, 1990) and continues to improve through early childhood (Carlson, 2005). Three- to four-year-old children typically perform poorly on go/no-go tasks by responding to no-go trials when they should not (Diamond, 2002). However, by 4.5 years of age, children’s performance

improves. There is evidence that inhibitory abilities in preschool are predictive of similar abilities in later life. In a study by Eigsti and colleagues (2006), children were assessed with a delay-of-gratification task at age 4, and then again 10 or more years later with a go/no-go task. Children who were able to divert their attention away from a tempting prize at age 4 had faster, and just as accurate, responses on the go trials 10+ years later.

However, children have, by no means, reached adult mastery on inhibitory functioning by the end of preschool. For instance, 4- and 5-year-old children have more difficulty than older children at preventing the intrusion of irrelevant information during recall tasks, an aspect of interference control (Bjorklund & Harnishfeger, 1995). Children also improve in their ability to suppress a dominant response and instead perform a less-dominant response up to age 6.5, as demonstrated by improved performance on a simpler version of the HTKS task (version with only “head” and “toes” commands) (Ponitz et al., 2008).

There seems to be another growth in inhibitory control sometime between the periods of 6-8 years and 9-12 years, where older children have demonstrated superior performance on assessments of inhibitory control up until early adolescence (around 12 years of age), when no further improvement is noted (Brocki & Bohlin, 2004; Levin et al., 1991; Williams, Ponesse, Schachar, Logan, & Tannock, 1999). Still others have noted that some forms of inhibitory control reach adult-like maturity earlier. In a study examining interference control, children were required to push a button on either the left or right side of a keyboard in correspondence with the direction a fish is facing on a computer screen (Fan, McCandliss, Sommer, Raz, & Posner, 2002). On some trials, the



fish is presented alone on the screen, and on other trials, flankers are present (distracter fish facing in opposite direction of the target fish). Flanker trials are difficult because children must attend to the direction of the target fish only and ignore the direction of the distracter fish. Children's ability to selectively attend to the target stimulus in the presence of conflicting information improves greatly between 4 to 7 years, and seems to remain fairly consistent after age 7 (Rueda, Posner, & Rothbart, 2005). Taken together, these studies seem to indicate that some forms of inhibitory control reach adult-like maturity before early adolescence, while others do not.

### **Cognitive Flexibility or Set Shifting**

The next component of EF, cognitive flexibility or set shifting, takes place in two phases. First, children form an initial rule where a certain stimulus (i.e. color of a card) requires a certain response (sort the card into appropriate color pile). Children must selectively attend to this piece of information (Garon et al., 2008). Then, demands of the task require that a new rule be held in mind (i.e. sort by shape of card instead), and the old rule must now be suppressed (ignore card color). Thus, a "shift" must take place. Most tasks require both a mental shift of attention and inhibition of an old (no longer relevant) response for task success (Garon et al., 2008).

The Dimensional Change Card Sort (DCCS; Frye, Zelazo, & Palfai, 1995), which involves sorting first by color and then by shape as prompted by the experimenter, is a popular task for assessing cognitive flexibility, especially the ability to shift attention (Garon et al., 2008). Diamond (2002) describes it as "perhaps the simplest possible test of task switching" (p. 475). Children are explicitly told the old and new sorting rules in this

task, and are reminded of them throughout the task. Typically, 3-year-olds find it difficult to sort by the new rule, despite adult prompting, but 4- and 5- year-olds are able to do so (Carlson, 2005; Carlson & Moses, 2001). It is likely that younger children continue sorting on the old rule despite feedback because of the cognitive conflict the task presents (Garon et al., 2008). Younger children perseverate on the card feature that was correct with the original sorting rule (color), and find it difficult to switch focus and attend to a different feature of the cards for the new rule (shape).

Though performance on the DCCS improves dramatically by the end of preschool, even older children and adults still fixate somewhat on the original sorting rule. This is evident in their slowed reaction times during the second block of trials when participants are required to sort on the new rule (Diamond, 2002; Diamond & Kirkham, 2001; Roger & Monsell, 1995). Zelazo (2006) developed an advanced version of the DCCS validated for use with children up to age 7. In this version, a third block of trials is added where both sorting rules are relevant throughout the trials. The stimulus cards are the same as described above, except that some cards now have a star on the corner, which indicates that the card should be sorted by color. Cards without stars are to be sorted by shape. There are fewer trials with star cards compared to non-star cards during this third block, and they are presented randomly. Five- and six-year-olds find it difficult to switch between sorting rules when a star card appears (Best & Miller, 2010; Carlson, 2005).

### **Planning Skills**

Planning is the last EF component to be examined in the current study. Planning requires working memory skills, and involves forethought, or the ability to think ahead

and envision possible solutions to a current problem (Barkley, 1997). The Tower of Hanoi (TOH; Simon, 1975) and Tower of London (TOL; Shallice, 1982) are two popular options for measuring planning skills (Barkley, 1997). They involve the participant moving rings or balls around on pegs in as few moves as possible to match the experimenter's ring/ball and peg set. These tasks increase in difficulty with the minimum number of moves required for completion. Success on preschool versions of the TOH (Carlson, Moses, & Claxton, 2004) improves greatly between the ages of 3 to 5 (Atance & Jackson, 2009). Planning again improves during mid to later childhood, as children demonstrate success with more difficult three- to five-move TOL problems (Levin et al., 1991; Luciana & Nelson, 1998). Luciana and Nelson found that children were successful at three-move TOL problems by middle childhood, but success on more complex four- or five-move problems was not reached until later in childhood. Similarly, Levin and colleagues found that planning on the TOL did not improve much from their youngest (7-8 years) to their middle age group (9-12 years). However, the oldest age group (13-15 years) showed significantly better planning skills than the youngest group on four-move TOL problems.

Despite all of the typical developmental trends just described, there is great inter-individual variability in the development of EF. Individual differences between children are related to different patterns of EF. This notion will be examined in more detail for children with externalizing behavior problems and/or children who speak more than one language.

## **Executive Functioning for Children with Externalizing Behavior Problems**

Unfortunately, everyday in the classroom, teachers are confronted with at least a few children who cannot seem to sit still, whose attention flits quickly from one activity or topic to another, and who act impulsively or even aggressively in interactions with peers. If these behaviors occur on a regular basis, these children are said to exhibit externalizing behavior problems. Such children are at risk for a future diagnosis of a disorder such as ADHD (Campbell, 2002). Young children with externalizing behavior problems often demonstrate impairments in EF, ranging from mild to severe (Barkley, 1997). They have difficulty regulating their own behavior and inhibiting impulses (Barkley, 1997), which may prompt caregivers to take the child to a mental health professional (Campbell, 2002). The severity of EF deficits varies with the severity of the behaviors, where children with a more severe deficit in a particular EF are likely to display more severe behavior problems. However, impairment in one component of EF does not necessarily imply impairment to the same extent in all other components of EF (Barkley, 1997). For instance, a child may have difficulty with sustaining attention and attentional control, but may not exhibit a deficit in working memory. Each component is affected differentially, with inter-individual variability in degree of impairment. Before further exploring the patterns of EF for these children, it is important to clarify what is meant here by “externalizing behavior problems” or “at-risk for ADHD.”

Extensive research on children has identified two broad categories of problem behavior: *internalizing* behaviors include extreme social withdraw, shyness, and anxiety around others, while *externalizing* behaviors include high frequency of aggression,

fighting, over-activity, inattention, and failure to follow instructions (Campbell, 2002). Externalizing behaviors are more noticeable by adults, as these behaviors typically cause disruption in the social environment through negative peer-to-peer and adult-to-child interactions. When these behavior patterns include over-activity, inability to sit still or “fidgetiness,” poor impulse control, impatience with peers, and/or difficulty with sustained attention, the child is at-risk for a diagnosis of ADHD (Campbell, 2002). Externalizing problems are also more likely than internalizing problems to endure over time (Campbell, 2002). Therefore, the current study focuses on *externalizing* problems as they relate to EF. The research summarized below is a small window into the literature on behavior problems and EF. Several of these studies used children already diagnosed with ADHD. Fewer studies have examined community samples of undiagnosed children at-risk for ADHD (Berlin & Bohlin, 2002). The reader will see that more consistent deficits are found for inhibitory control, with less consistent findings for set shifting and planning.

Children with externalizing behavior problems consistently demonstrate difficulties with several aspects of inhibitory control. Children diagnosed with ADHD perform poorly compared to controls on studies examining delay-of-gratification, a measure of behavioral inhibition (Campbell, Pierce, March, Ewing, & Szumowski, 1994; Rapport, Tucker, DuPaul, Merlo, & Stoner, 1986). ADHD children also struggle with inhibiting a prepotent response, as demonstrated by worse performance than controls on the no-go trials of the go/no-go task (Iaboni, Douglas, & Baker, 1995; Shue & Douglas, 1992; Trommer, Hoepfner, Lorber, & Armstrong, 1989). Similarly, there is evidence for

impairment in the non-clinical population of children with behavior problems, where those rated as having more behavior problems perform worse on no-go trials (Berlin & Bohlin, 2002; Brophy et al., 2002). In a study by Raaijmakers et al. (2008), a non-clinical sample of children with aggressive behavior demonstrated poorer inhibitory functioning compared to controls across several tasks. This finding held even after controlling for attention problems. The authors interpreted this to mean that aggression is related to deficits in EF (inhibition specifically), above and beyond ADHD symptoms.

Additionally, children with ADHD tend to perform worse than controls on measures of interference control, such as the Stroop task (Barkley, Grodzinsky, & DuPaul, 1992; Leung & Connolly, 1996; Pennington, Grossier, & Welsh, 1993; Seidman et al., 1996). Other studies using the Day-Night Stroop-like task (Gerstadt et al., 1994) have also found poorer performance for children with behavior problems (Berlin & Bohlin, 2002; Berlin, Bohlin, Nyberg, & Janols, 2004), indicating that these children have difficulty with suppressing a dominant response and instead replacing it with a less dominant response (e.g. say “night” to a picture of a sun).

Findings are not as straightforward for the relationship between behavior problems and cognitive flexibility (set shifting). Several studies have found that children with externalizing problems are impaired on measures of cognitive flexibility compared to controls. This is not surprising, as this EF component relies on both working memory *and* inhibitory control (Garon et al., 2008). A recent study used a modified version of the DCCS with ADHD children and controls, found that the ADHD group made more errors once the sorting rule changed and they had slower reaction times (Mulas et al., 2006).

Barkley et al. (1992) reviewed 13 studies that examined performance on the Wisconsin Card Sort Test (WCST) for ADHD children. This task involves sorting cards by one rule first for several trials, and then deducing when the sorting rule has changed through feedback about response correctness (Heaton, 1981). It is more complicated than the DCCS in that there are more dimensions to attend to when sorting the cards. The WCST has been used often as a measure of cognitive flexibility (Barkley et al., 1992; Winsler, Abar, Feder, Schunn, & Rubio, 2007). ADHD children performed worse on this task in eight of the 13 studies reviewed by Barkley et al. The null findings in the other five studies might be due to small sample sizes and wide age ranges (Barkley et al., 1992). In a more recent study by Winsler and colleagues, a group of children with an autism spectrum disorder (ASD) performed significantly worse on the WCST than controls, but the performance of an ADHD group was not significantly different from either the ASD or control groups. Barkley (1997) noted that both typical *and* ADHD children improved with age on the WCST. This provides evidence for a delay in EF for children with behavior problems, instead of a qualitative difference from peers.

Findings regarding the planning abilities of children with behavior problems are also inconsistent across studies. Some have found that children with ADHD perform worse than controls on the Tower of London (TOL) and/or Tower of Hanoi (TOH; Kopecky, Chang, Klorman, Thatcher, & Borgstedt, 2005; Pennington et al., 1993; Weyandt & Willis, 1994), two classic assessments of planning. However, in a non-clinical sample of “hard to manage” 7-year-olds, Brophy et al. (2002) found no differences between this group and controls on a TOL-like task. This null finding could

be due to task format (conducted on a computer instead of manually), or to inadequate test-retest reliability with the TOL (Bishop, Aadmott-Leeper, Creswell, McGurk, & Skuse, 2001). Another potential explanation is that the non-clinical sample (also examined at age 4, when differences in planning were found) simply improved over time in their planning skills relative to control children. Kopecky and colleagues (2005) found that ADHD children diagnosed with both hyperactivity-impulsivity *and* attention problems performed worse on the TOH than those with attention problems only.

In a review of 18 studies that examined EF in children with ADHD (Pennington & Ozonoff, 1996), only three did not find significant differences in performance between ADHD children and controls (they included tasks of cognitive flexibility, such as the WCST). Further, no study indicated that the ADHD group performed *better* than controls on any EF task. Pennington and Ozonoff concluded that performance on behavioral inhibition measures such as the go/no-go task seems to be consistently impaired for ADHD children, while impairment is less consistent for other aspects of EF (e.g. cognitive flexibility). In a recent meta-analysis of 83 studies on EF and ADHD (Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005) – some with diagnosed samples and others with community (undiagnosed, but at-risk) samples – all studies showed impairments in EF for participants with behavior problems. Effects sizes were moderate (.46-.69), with the strongest effects for deficits in planning and behavioral inhibition. I will now turn to the research on EF and its relationship with a very different construct: the acquisition of more than one language in childhood.



## **Executive Functioning for Children Who Speak More Than One Language**

A growing body of research is demonstrating that children who are balanced bilinguals have an advantage in EF over monolinguals, known as the “bilingual advantage” (Bialystok, 2001). In the context of the current paper, “balanced bilingual” refers to an individual who can use both his/her native (L1) and second language (L2) about equally well across multiple contexts (Bialystok, 2001). There is mounting evidence that speakers of two languages constantly have both languages mentally activated (Guttentag, Haith, Goodman, & Hauch, 1984). It is thought that the continual challenges these children face in keeping two languages framed in mind, knowing when to switch from L1 to L2, and inhibiting one while using the other provides ample opportunities to practice selective attention and inhibition (Bialystok, 2001).

For instance, when talking with an English-speaking preschool teacher, young Spanish-English bilinguals must inhibit their Spanish vocabulary and instead attend to the current situational context and produce the English vocabulary they have in their repertoire. The two languages the young bilingual holds in mind compete for attention, and the child must choose which to attend to, and which to inhibit, in a given situation (Martin-Rhee & Bialystok, 2008). Green (1998) proposed that the mechanisms used to suppress the unneeded language are the same as those used during typical inhibitory functioning and selective attention. This “practice” at mentally inhibiting one of the languages while attending to the other seems to generalize to other EF components (Bialystok, 2001), resulting in the bilingual advantage. Research evidence will be

examined for each EF component, examining inhibitory control first, followed by cognitive flexibility, and finally a brief discussion on planning skills.

Interference control, a form of inhibitory control, is one of two EF components (along with cognitive flexibility) that most consistently demonstrates a bilingual advantage. Bilinguals and monolinguals appear to perform similarly on inhibitory tasks that are more response-based in nature and that do not involve much conflict or distraction (Bialystok & Martin, 2004). Examples include the classic go/no-go paradigm, where a child must simply suppress a dominant (prepotent) response, and the Day-Night task, where a child must choose between competing responses (Martin-Rhee & Bialystok, 2008). However, when perceptual or semantic conflict becomes part of the task and attentional/cognitive control becomes necessary for successful task performance, bilingual children seem to have an advantage, indicating superior interference control compared to monolinguals.

The Simon task is an assessment frequently used in research on EF with bilinguals (Baker, Kovelman, Bialystok, & Petitto, 2003; Bialystok, Craik, Klein, & Viswanathan, 2004; Martin-Rhee & Bialystok, 2008). This task presents children with perceptual conflict that needs to be resolved in order to be successful (Martin-Rhee & Bialystok, 2008), involving interference control and selective attention. The participant pushes a button on the left when a certain stimulus is presented (i.e. a blue square), and a button on the right when a different stimulus is presented (i.e. a red square). Congruent trials occur when the stimulus is presented on the same side of the computer screen as the button to be pushed. Incongruent trials occur when the stimulus is presented on the

opposite side of the screen from the correct button (Simon, 1969). Therefore, the incongruent trials present conflicting information – the type of stimulus and physical location of the stimulus – and the child must choose which to attend to. If the hypothesis of mental inhibitory practice and interference control for bilinguals is correct, bilinguals should show an advantage over monolinguals for the incongruent trials of the Simon task, but not necessarily for the congruent trials (Bialystok et al., 2004). However, in three studies by Martin-Rhee and Bialystok (2008), bilingual children ranging in age from four to eight years old performed better than monolinguals on both the incongruent *and* congruent trials of the Simon task. Researchers have interpreted this to mean that interference control is present on both trial types because the trials are presented randomly, so the child never knows which type is coming next (Lu & Proctor, 2001; Martin-Rhee & Bialystok, 2008). Similar results were found in another study where both younger (30-54 years) and older (60-88 years) adult balanced bilinguals outperformed monolinguals on incongruent and congruent trials, indicating less cognitive interference during the task for these individuals (Bialystok et al., 2004; Study 1). It is important to note, however, that the size of the effect (i.e. the advantage for bilinguals) was greater for incongruent trials.

On the other hand, bilinguals do *not* appear to have an advantage in inhibitory control when no perceptual (or semantic) conflict is present, or when required to simply suppress a response (Bialystok & Martin, 2004; Bialystok & Viswanathan, 2009; Carlson & Meltzoff, 2008; Emmorey, Luk, Pyers, & Bialystok, 2008; Martin-Rhee & Bialystok, 2008). For instance, using the Day-Night Stroop-like task, Martin-Rhee and Bialystok

(2008) found no difference in performance between 4-year-old bilinguals and monolinguals. Additionally, Emmorey and colleagues (2008) found that bilinguals did *not* demonstrate fewer commission errors on the no-go trials of the go/no-go task. Commission errors on these trials (child pushes the button to the no-go stimulus) are thought to be a direct measure of behavioral inhibition (Berlin & Bohlin, 2002). While bilinguals do have experience with resolving conflict between their two competing languages, they do not have additional experience compared to monolinguals at simply stopping a response or performing a less-dominant response in place of a dominant one. The equivalent of this would be a bilingual child simply not speaking, as opposed to choosing between English and Spanish when speaking (Martin-Rhee & Bialystok, 2008).

In another study, Carlson and Meltzoff (2008) used a sample of kindergartners and presented them with several tasks involving some kind of conflict between stimuli (i.e. Dimensional Change Card Sort, Simon Says task, Flanker Task) and other tasks involving delay-of-gratification (food delay, gift delay, standing like a statue). They found that simultaneous bilinguals (i.e. individuals who learn both languages from birth, as opposed to learning one later in life) had higher conflict composite scores than monolinguals, indicating superior performance on conflict tasks. However, the groups did not differ on delay scores, again indicating no bilingual advantage for simply inhibiting a motor response. (Carlson & Meltzoff also included a group that was not fully bilingual, which is discussed in further detail below).

Cognitive flexibility, or set shifting, is the other EF component that shows a consistent advantage for bilinguals in prior research. In young children, cognitive

flexibility is often assessed with the Dimensional Change Card Sort (DCCS; Frye et al., 1995). This task taps into many EFs, including working memory (holding the current sorting rule in mind), set shifting (shifting to a new sorting rule and holding this new rule in mind), and inhibition (cognitively inhibiting the old sorting rule and/or behaviorally inhibiting the actual motor responses that are no longer correct). Bialystok and Martin (2004) found superior performance of 4- and 5-year-old bilinguals on a variation of the DCCS. Participants first sorted cards by color, then by color and shape, then by color and object, and finally by function and location. Bilinguals and monolinguals performed similarly with the first and last sorting rules, but bilinguals performed better with the second and third rules. The first sorting rule was easy for all participants, so bilinguals and monolinguals performed similarly. The last sorting rule (sorting by function, then location) was too difficult, so all participants performed equally poorly. But the second and third sorting rules were just difficult enough where the advanced inhibition and set shifting of bilinguals led to greater success (Bialystok & Martin, 2004).

This finding is robust across a variety of language groups. Bialystok and Martin (2004; studies 1 and 3) noted the bilingual advantage on the DCCS for Chinese-English bilinguals. In Study 2, they found the same pattern of results for French-English bilinguals. In a more recent study, Okanda, Moriguchi, and Itakura (2010) examined DCCS performance for a sample of preschool-age Japanese-French bilinguals. The bilinguals out-performed monolinguals when the two groups were matched on verbal ability. In an earlier study with the DCCS, Bialystok (1999) noted that 4-year-old bilinguals performed at the level of 5-year-old monolinguals. This supports the notion

that bilinguals are *advanced* in EF development, as opposed to qualitatively different from monolingual peers.

There is additional evidence for bilinguals' superior ability to focus and attend selectively to specific pieces of information. Monolingual and bilingual 6-year-olds performed similarly on an embedded figures test, where they are required to locate a shape hidden inside a larger picture, such as locating a small picture of a house embedded in a larger picture of a robot (Bialystok & Shapero, 2005). Bialystok and Shapero reasoned that this similar performance between groups is because the embedded figures task does not present any perceptual conflict. However, in this same study, children also completed an ambiguous figures task. With this task, an ambiguous image can be seen in one of two possible ways (i.e. an image that is both a vase and a pair of faces). Children had to identify one image first, and then were prompted to find the second image in the picture. Bilinguals outperformed monolinguals on this task. Bialystok and Shapero reasoned that this task presents perceptual conflict because only one image can be perceptually processed at a time (whereas both images could be seen at the same time with the embedded figures task). Bilinguals were better at managing this conflict. It is interesting to note that the bilinguals in this study spoke English as one language, and a variety of other possibilities as their second language, including French, Korean, Chinese, Spanish, Russian, Portuguese, Hebrew, Hungarian, and Kurdish (Bialystok & Shapero, 2005). Overall, bilinguals demonstrate a clear advantage over monolinguals on tasks of cognitive flexibility.

Unlike with interference control and cognitive flexibility, there has not been much research that specifically examines the planning abilities of bilinguals. The few studies that did examine this found some evidence for a bilingual advantage. For instance, Festman, Rodriguez-Fornells, and Münte (2010) compared performance on the TOH for adult bilinguals with greater proficiency in their two languages to that of bilinguals with less proficiency (17 to 45 years old). The more proficient adults performed better on the TOH than the less proficient adults in terms of fewer errors and fewer moves to achieve success. In one unpublished study, bilingual children performed superior to monolingual children (age unavailable) on the Tower of London (TOL; Moreno, Case, Schellenberg, & Bialystok, 2010). Surprisingly, no published studies could be found that examined planning skills of bilingual *children*. Perhaps successful completion on tasks like the TOH tap mostly into working memory (holding multiple possible moves or solutions in mind) and behavioral inhibition (refraining from responding too quickly so that the fewest number of moves can be used), two EF components that do not seem to demonstrate a clear bilingual advantage (Bialystok et al., 2004; Martin-Rhee & Bialystok, 2008). It has been proposed that the Trail-Making Test (Reitan, 1971) requires planning skills, especially Trail B. In Trails A (part 1) of this task, a participant must connect dots linking numbers in sequential order by starting at the dot for number 1, and then draw a line from this dot to the dot for number 2, then to number 3, and so forth. In Trails B (part 2), the participant connects dots for alternating letters and numbers by starting at the dot for number 1, then drawing a line to letter A, then drawing a line to number 2, then to

letter B, and so forth. Bilingual five- and six-year-olds have demonstrated superior performance over monolinguals on this task (Bialystok, 2010).

While balanced bilinguals clearly show an advantage in certain EF components compared to monolinguals, very little is known about the functioning patterns for children who are still acquiring a second language (not fully proficient in both languages). These children are in the process of becoming fully proficient in two languages, and may one day be considered balanced bilinguals, but they are not there yet. For instance, in the United States, non-native speakers of English make up a large portion of the population of children under age 18. As recently as 2007-2008, more than 5.3 million children in United States public schools (10.7% of school-age population) were considered to be English Language Learners (NCELA, 2010).

Before discussing the few studies that have examined EF for such children, it is important to clarify additional issues related to bilingualism and second language learning, as terms are often unclear, and definitions are not consistent across studies (Baker, 2006). One important issue is the balance of the two languages for an individual; one language is usually dominant (Baker, 2006). As noted above, “balanced bilingual” refers to an individual who can use both his/her native (L1) and second (L2) language about equally well across multiple contexts (Bialystok, 2001), and/or is considered to be at or above age-level competence in each language (Baker & Jones, 1998). Another important issue is age of acquisition of L2. Individuals who acquire both languages from birth are referred to as simultaneous bilinguals, while those who acquire L2 after the toddler years are sequential bilinguals (Baker, 2006). Sequential bilinguals who are in the



process of acquiring L2 (i.e. an immigrant child who has been learning English for the past three years) are often referred to as dual language learners (DLLs).

When studying DLLs, it is important to distinguish between *ability* with the second language and actual everyday *use* of that language (Baker, 2006). Language ability (or proficiency) can be thought of on a continuum, with some DLLs only having receptive skills in the second language (L2), while others read and write mostly in L2, and still others read and write in both the native language (L1) and L2 (i.e. balanced bilinguals). For the current paper, it is also important to consider whether the bilingual advantage in EF is present for bilinguals *and* DLLs, or bilinguals only. In other words, how might a child's *degree of bilingualism* (the balance between the two languages) relate to EF? Based on the limited prior findings that are available, the current paper proposes that DLLs, who are expected to have less ability in or use of L2, would show less of an EF advantage compared to balanced bilinguals who are highly proficient in and/or frequently use both languages. Unfortunately, most of the research in this area draws from the population of balanced bilinguals; few studies have specifically examined DLLs (Baker, 2006; Bialystok, 2001). I will now summarize three studies that have examined DLLs and what this might mean for the relationship between language *proficiency* and EF.

Carlson and Meltzoff (2008) compared the performance of three groups of kindergarten children – balanced bilinguals, DLLs, and monolinguals – on a variety of EF tasks. They defined bilingual as having exposure to Spanish and English from birth. Their DLL group was comprised of English monolinguals who started attending a

language immersion program in kindergarten. They were taught in either Spanish or Japanese for approximately three hours each school day (and English for the rest of the day). Parents confirmed that children were English monolinguals at the start of kindergarten, and children had been in the immersion program for about six months when the study began. Control children were English monolinguals (parent report) with only 30 minutes maximum per week of exposure to another language (or no exposure). Children completed a variety of tasks designed to tap into either delay-of-gratification or resolution of conflicting information. Results indicated no differences on the EF tasks between the DLL and control groups when examined at the individual task level or as one overall EF composite score.

Carlson and Meltzoff (2008) also conducted a factor analysis on all EF outcome variables, resulting in two factors labeled *conflict* and *delay*. They then created two composite scores, one for the tasks that required resolving conflict and one for those that required delay-of-gratification. There were no group differences between bilinguals, DLLS, and monolinguals on the delay score. Bilinguals outperformed the other groups on conflict, but there was no difference between DLLS and monolinguals. Exposure to a second language for six months did not seem to provide the DLLs with the traditional “bilingual advantage” in EF. It may be that these children had not reached the threshold needed to see the advantage. In other words, it might be that more exposure to, and presumably greater proficiency with, the second language is required to have an influence on EF (Carlson & Meltzoff, 2008). An alternative possibility is that exposure to two languages must begin at younger ages (i.e. younger than 5) to see an advantage in EF. My

study improves on this design by recruiting children with at least six months exposure to a second language (though most DLLs in the current study had much more than six months exposure), but who are still not considered to be fully balanced bilinguals based on parent report of proficiency and actual language use, and direct assessment of receptive and expressive language skills.

Bialystok and Majumder (1998) examined three groups of third graders ranging in age from 7 to 9 years. Groups included monolingual English children, French-English balanced bilinguals who received schooling in French, but had French and English in the home and community, and Bengali-English DLLs who received schooling in English, but had parents that spoke Bengali at home. This last group had limited (if any) formal instruction in Bengali, so they were considered to be stronger in English. Tasks included a metalinguistic task (judging grammar of sentences), a block design task, a water level task, and a final task assessing children's understanding of proportions. While balanced bilinguals performed better than the other groups on part of the metalinguistic task and on the block design task, there were no differences between DLLs and monolinguals. It appears that fully balanced bilinguals had some advantage over the other groups in attentional control skills (grammar and block design tasks), but DLLs did not demonstrate an advantage.

In a final study by Luciana and Nelson (2002), participants completed a non-verbal working memory task (a search task), a TOL-like task (*Stockings of Cambridge*; Shallice, 1982), and a set-shifting task (in addition to other tasks not relevant here). One sample of participants in their study included both native and non-native English

speaking children in grades 1 to 6. It is unclear how the authors determined that the children were non-native speakers of English. They did not mention English proficiency level or age of acquisition for the non-native group (though non-native English speakers scored lower on a measure of English vocabulary, but scored similarly to native English speakers on a non-verbal measure of IQ). Results indicated no significant differences on any of the EF tasks for children who spoke English as a second language compared to native speakers (Luciana & Nelson, 2002). However, it is difficult to draw conclusions from this study, as very limited information was available regarding the language experience of the DLL group. Overall, these three studies tentatively suggest that there is an unclear threshold of proficiency in two languages that children must reach in order to see benefits in non-linguistic domains. In the current study, I will attempt to further examine how *degree* of bilingualism might relate to EF (discussed further below in Rationale section).

Not all researchers agree that the bilingual advantage is actually attributable to learning two languages. Morton and Harper (2007), for instance, examined performance on the Simon task for 17 bilingual and 17 monolingual children (6-7 years old) who were matched on ethnicity and SES. Parents indicated that monolingual children only understood and spoke English. Bilingual children attended school in French. Results indicated no language group differences in performance on the Simon task. They concluded that selective attention and inhibition were not influenced directly by bilingualism. They also pointed out that frequency of use of the two languages should be of more importance in future studies, as opposed to just language proficiency. Namazi

and Thordardottir (2010) looked at controlled attention and working memory for 4- to 6-year-old bilingual and monolingual children. They did not find evidence for a bilingual advantage on the Simon task. However, they did find better performance on the Simon task for children with more advanced visual working memory, regardless of language group. They concluded that the advantage reported for bilinguals on such tasks might be due to working memory, rather than to bilingualism specifically. Based on these conflicting findings, it is important to further examine the relationship between bilingualism and EF, particularly, the language use and proficiency in *both* languages, as well as age of acquisition of L2.

### **Bilingualism, Language Development, and Externalizing Behavior Problems**

Research has demonstrated a strong link between a child's language skills/development and behavior problems. Several prior studies have demonstrated that children with language disorders, such as specific language impairment, have limited competence during social interactions compared to typically developing peers (Beitchman et al., 1996; Fujiki, Brinton, & Todd, 1996; Rice, 1993; Rice, Sell, & Hgadley, 1991). But these findings are not limited to children with diagnosed language disorders; the language skills of typically developing children are also related to competence in social interactions (Fagan & Iglesias, 2000; Farmer, 1997; Hart, Olsen, Robinson, & Mandlco, 1997; Hazen & Black, 1989).

From this we can conclude that, for all children in general, language ability should be related to behavior problems (Fagan & Iglesias, 2000; Hart et al., 1997).

Indeed, Fagan and Iglesias (2000) found that children's communicative abilities were directly linked to behavior problems at the same point in time (Time 1), as well as *predictive* of behavior problems at a later point in time (Time 2), even after controlling for behavior at Time 1. More specifically, the authors found that children's conversational skills, and not the structural complexity of children's language, were predictive of Time 2 externalizing behavior problems. Many other studies have also indicated a negative relationship between language competence and behavior problems, where poor language/literacy skills are associated with more behavior problems (Benasich, Curtiss, & Tallal, 1993; Bub, McCartney, & Willett, 2007; Bulotsky-Shearer & Fantuzzo, 2011; LaParo & Pianta, 2000; Miles & Stipek, 2006; Spira & Fischel, 2005). It is unclear whether delays in language and literacy lead to externalizing behavior problems, or whether behavior problems lead to communication delays. It is most likely a transactional process at play, where language delays fuel behavior problems, which in turn limit the amount of educational opportunities for the child, thereby further disrupting language development.

A large body of research has also indicated that children with behavior problems are delayed in the internalization of private speech (see Winsler, 2009 for a review). Private speech (i.e. talking to oneself) is related to executive functioning, and is an important tool for self-regulation (Berk, 1992; Winsler, 2009). During early childhood, children make frequent use of private speech, which helps them monitor attention level, plan out actions, and problem-solve (Berk, 1992; Winsler, 2009). Studies examining private speech in both community (Winsler et al., 2000) and diagnosed samples (Winsler,

1998) find that children with behavior problems frequently use private speech, though they have delayed speech internalization compared to their typically developing, same-age peers. These findings help make the important distinction that children with behavior problems are delayed in functioning compared to same-age peers, as opposed to qualitatively different.

If advanced or simply age-appropriate language skills are related to fewer behavior problems for children, this leads to the question of how knowledge of *two* languages might be related to behavior problems and socio-emotional competence. Research examining the relationship between bilingualism and behavior problems is relatively new. Despite this, the findings of a few recent studies indicate that children who are categorized as bilingual or English language learners (ELLs) appear to have an advantage in socio-emotional competence, including fewer behavior problems, compared to their monolingual peers (De Feyter & Winsler, 2009; Hair, Halle, Terry-Humen, Lavelle, & Calkins, 2006; Han, 2010; Han & Huang, 2010). For instance, De Feyter and Winsler found that preschool children in immigrant families (many of whom speak a language other than English) demonstrated fewer behavior problems and stronger socio-emotional skills compared to non-immigrant peers.

Using data from the Early Childhood Longitudinal Study-Kindergarten cohort (ECLS-K), Hair and colleagues (2006) used cluster analysis to group children into various profiles in kindergarten. The analysis resulted in four profiles: comprehensive positive development, strengths in socio-emotional and health, risks in socio-emotional, and risks in health. Interestingly, those children with a profile of socio-emotional and

health strengths were also *least* likely to come from a family where English was the dominant language spoken. Hair et al. hypothesized that, for children from language-minority backgrounds, socio-emotional skills may be more important than other traditional academic measures in terms of school readiness.

Others using data from the ECLS-K have examined change over time in socio-emotional competence and behavior problems for different language groups (Han, 2010; Han & Huang, 2010). Han and Huang examined Asian children and non-Hispanic White children from kindergarten through fifth grade. At the start of kindergarten, there was no significant difference in behavior problems for the language groups. However, over time, children who were considered either balanced bilinguals or ELLs had smaller increases in behavior problems compared to monolingual English-speaking peers. Interestingly, children who were considered monolingual, but spoke a language other than English, demonstrated a greater increase in behavior problems over time compared to the other language groups. Han reported similar findings for Latino elementary school children, where balanced bilinguals and ELLs demonstrated fewer behavior problems than native English-speakers. In other words, these studies indicate that bilinguals and ELLs demonstrate better behavior than native English-speaking students, at least during elementary school. Additionally, the absence of behavior problems appears to be related to the acquisition of English among language minority students (Kim, Richard, & Winsler, 2012). Kim and colleagues found that native Spanish-speaking preschoolers in Miami who had fewer behavior problems were more likely than their peers with more behavior problems to become proficient in English by the end of kindergarten.



However, not all research on language minority children supports a positive link between language minority status and fewer behavior problems. Preciado, Horner, and Baker (2009) discussed the need for interventions to help improve reading skills, as well as behavior, for Latino elementary students. It is well known that native Spanish-speaking ELLs tend to struggle academically in the U.S. (August & Hakuta, 1998; Klingner & Artiles, 2006). According to Preciado et al. and others (McIntosh, Chard, Boland, & Horner, 2006; Sanford, 2006), students faced with academic difficulties often develop behavior problems as a way to avoid dealing with academic tasks. Due to their historically lower academic performance, Preciado et al. concluded that Latino ELLs are at particular risk for the eventual development of behavioral issues. In light of these findings, it is important to consider how behavior problems and cognitive functioning may be related for children who speak more than one language.

The current study will be the first to incorporate consideration of behavior problems into the bilingualism-EF relationship by exploring both mediation and moderation models. First, I examined whether EF mediates the relationship between bilingualism and externalizing behavior. Prior research indicates a strong positive relationship between bilingualism and EF, and it is believed that regular use of two languages *leads to* the EF advantages (Bialystok, 2001). There is also strong evidence for a relationship between EF and behavior problems, where children with EF difficulties demonstrate more externalizing behaviors (Brophy et al., 2002; Raaijmakers et al., 2008). Finally, there is some evidence that children who are fluent in multiple languages demonstrate fewer behavior problems and have stronger socio-emotional skills (De

Feyter & Winsler, 2009; Han, 2010; Han & Huang, 2010). Therefore, I examined whether the relationship between bilingualism and externalizing behavior exists *because of* the association each construct has with EF.

I also examined whether language group moderated the relationship between EF and behavior. Research has indicated that the development of children's language, particularly private speech, reorganizes prefrontal and other neurological systems that facilitate self-regulation (Luria, 1961; Vygotsky, 1986; Winsler, Fernyhough, & Montero, 2009). It is possible then, that the development of *two* languages might reorganize prefrontal systems in additional ways, such that EF may relate differently to other areas of functioning for these children.

### **Rationale for the Current Study**

The current dissertation contributes to the literature in several ways. First, I am focusing on an age range (5 to 7 years old) that has not received as much attention in terms of research on EF. In the past decade, a large body of research has examined EF during the preschool period, with some neglect of the early elementary school years. Second, I am seeking to replicate the bilingual advantage in EF found in prior studies (Baker et al., 2003; Bialystok, 1999; Bialystok et al., 2004; Bialystok & Martin, 2004; Martin-Rhee & Bialystok, 2008; Okanda et al., 2010), particularly because findings are mixed, and not all studies found support for the bilingual advantage (Morton & Harper, 2007; Namazi & Thordardottir, 2010). However, I am also building on this prior research by adding additional tasks not examined (or examined very little) with bilingual children. Specifically, I included an inhibitory control task that has demands similar to that of the

Day-Night Stroop task, but requires more than simple response suppression, known as the Head-Toes-Knees-Shoulders task (HTKS; Ponitz et al., 2009). Like the Day-Night Stroop, the HTKS requires a child to inhibit a habitual response in order to perform an opposing, less-dominant response (explained in more detail in Measures section). In addition, the HTKS is more response-based in nature and does not present conflicting information to the child (no interference control required). As already described, a bilingual advantage has not been found for tasks like go/no-go and Day-Night Stroop (Emmorey et al., 2008; Martin-Rhee & Bialystok, 2008). It is important to determine whether a similar pattern exists for a task like the HTKS that requires more overt motor movement, but still does not require cognitive/attentional control like the Simon and original Stroop tasks do. Based on prior research indicating that young bilingual/DLL children have better socio-emotional skills and fewer behavior problems compared to monolinguals (De Feyter & Winsler, 2009; Han, 2010; Han & Huang, 2010), it is possible that they might perform superior to monolinguals on this type of task, and the advantage may be due to something other than the documented “bilingual advantage” (as there is not a reason to expect an advantage in this type of inhibitory control that is attributable to dual language use).

I also included the Tower of London (TOL; Shallice, 1982) with my battery of tasks. The TOL is popular for the assessment of planning abilities in children (Fernyhough & Fradley, 2005; Levin et al., 1991; Lidstone, Meins, & Fernyhough, 2010; Luciana & Nelson, 1998). However, limited prior research with bilinguals does not yet allow for a conclusion for or against a bilingual advantage in planning skills. More

research with bilinguals that involves complex tasks that engage several aspects of EF, such as the TOL, is necessary in order to further theoretical understanding of the relationship between bilingualism and cognitive development.

Another major contribution of my study is the way in which I operationalize a child's language status. I examined a child's language status as both a categorical variable (monolingual vs. DLL vs. bilingual), as well as a continuous variable (*degree* of bilingualism). Carlson and Meltzoff (2008) noted that there is not an agreed upon standard in the field for what it means to be "fully bilingual." Some have also noted that researchers do agree that it is better to think of bilingualism as continuous, or *how* bilingual a given individual is (Carlson & Meltzoff, 2008; de Villiers & de Villiers, 2010). Despite this, most research examining cognitive functioning has failed to do so, and treats bilingualism as a yes/no variable (balanced bilinguals vs. monolinguals). In the current study, several pieces of information were used to create a continuum for bilingualism. Children received a score for how balanced their two languages are (where one end of the spectrum indicated less balanced, or less bilingual, and the other end indicated more balanced, or more bilingual), and this score was used in analyses. Further, many prior studies examining bilingualism rely on a receptive vocabulary measure to assess children's language (de Villiers & de Villiers, 2010), often only in one language. It is important to consider *expressive* as well as receptive language skills, and to measure skill in *both* languages. For these reasons, I included a measure of expressive language abilities, and directly measured receptive and expressive language skills in both languages for bilinguals. I am also attempting to clearly and carefully define

bilingualism, and what it means to be a balanced bilingual vs. a DLL, something that is often overlooked or not well explained in prior studies.

Lastly, I am bringing behavior problems into the equation. As already discussed, bilinguals show advantages in tasks requiring cognitive control; however, due to their documented superior socio-emotional skills in early childhood, it is also possible that bilinguals might show advantages in tasks that are more response-based in nature.

Therefore, I examined potential mediating and moderating models for the relationships between bilingualism, EF, and externalizing behavior problems. Research questions and hypotheses included the following:

1. Are group differences in EF (three kinds inhibitory control, cognitive flexibility, planning skills, and overall parental report) evident for monolinguals, DLLs, and balanced bilinguals? I expected that group differences in some EF components would be evident. Balanced bilinguals would outperform both monolinguals and DLLs on measures of interference control (Simon task) and cognitive flexibility (Dimensional Change Card Sort; DCCS), but monolinguals and DLLs would not differ from one another. No language group differences were expected for simpler forms of inhibitory control (go/no-go and HTKS tasks). Due to limited prior research, there are no specific hypotheses regarding group differences in planning skills (Tower of London; TOL).
2. Within the group of children that speak a second language, do the measured EF components (three types of inhibitory control, cognitive flexibility, planning skills) vary in relation to degree of bilingualism? Scores for interference control (Simon)

and cognitive flexibility (DCCS) were expected to vary in a positive linear fashion with degree of bilingualism, where children who were “more balanced” would score higher on these measures of EF. Degree of bilingualism would be unrelated to simpler, more response-based forms of inhibitory control (go/no-go; HTKS).

3. Are group differences in behavior problems evident for monolinguals, DLLs, and balanced bilinguals? I expected group differences, where balanced bilinguals and DLLS would demonstrate fewer behavior problems (as reported by parents) than monolinguals.
4. Excluding monolinguals, are behavior problems related to degree of bilingualism? Due to limited prior research, I did not have specific hypotheses about the relationship between degree of bilingualism and behavior problems.
5. Do behavior problems vary in relation to any of the measured EF components (three types of inhibitory control, cognitive flexibility, planning skills)? I hypothesized that *all* EF components would have a *negative* relationship with behavior problems, where children with more problems would demonstrate greater deficits across all measured aspects of EF.
6. A series of mediated and moderated multiple regressions were conducted to examine two theoretical relationships between EF, degree of bilingualism, and behavior problems. First, I examined whether any EF component (i.e., three types of inhibitory control, cognitive flexibility, and planning skills) *mediates* the relationship between language group (monolinguals, DLLs, balanced bilinguals) and behavior problems. As already discussed, there are established relationships

between bilingualism and EF, and between EF and behavior, as well as some evidence linking bilingualism and behavior. Therefore, I propose that knowledge of two languages leads to enhanced EF, which in turn, leads to fewer externalizing behavior problems. Despite the fact that bilingualism is not supposed to lead to advantages in simpler forms of inhibition, these were also tested in the mediation models (go/no-go, HTKS), as research on the interplay between EF, bilingualism, and behavior has not been done before. Next, I examined whether language group *moderates* the relationship between EF and behavior problems. I expected that language group would moderate the relationship between EF (inhibitory control, cognitive flexibility, planning) and behavior, where the relationship expected between these variables (as EF improves, behavior problems decrease) would be stronger for children who are DLLs or balanced bilinguals, but weaker for monolinguals. I based this expectation on the idea that full facility with two languages might reorganize the brain and EF systems for bilingual children, similar to how the development and then use of language (L1) for self-regulation in the form of private speech is thought to reorganize prefrontal and other neurological systems to allow for self-regulation and EF to occur (Luria, 1961; Vygotsky, 1986; Winsler, Fernyhough, & Montero, 2009). If this is the case, then EF may in fact relate differently to other areas of functioning, such as problem behavior, for bilingual vs. monolingual children.

## METHOD

### Participants

Participants included 79 children ranging in age from five to seven years (two participants were 4 years old, and one was 8 years old, within one month of their 5<sup>th</sup> and 7<sup>th</sup> birthdays, respectively). Thirty-three participants (41.77%) were monolingual English speakers and the remaining 46 (58.23%) were bilingual Spanish-English speakers, including both balanced bilinguals ( $N = 17$ ) and Dual Language Learners (DLLs;  $N = 29$ ). [Four additional participants were excluded from all analyses for the following reasons: too much missing data ( $N = 1$ ); language scores so low in both English and Spanish indicating a likely serious language delay or disorder ( $N = 2$ ); and too little Spanish exposure to be considered DLL (4 months at an immersion school), but too much to be classified as monolingual ( $N = 1$ )].

Within the DLL group, most ( $N = 26$ , 89.66%) were classified as English dominant, two were Spanish dominant, and one could not be classified due to inconsistent language data (stronger Spanish receptive skills, but stronger English expressive skills). This characteristic of the DLL group differs from the original study goal, which planned to have a sample of mostly Spanish-dominant DLLs (still learning English). The resulting sample of English-dominant DLLs was due to sampling limitations; there were limited locations/outlets for recruiting bilingual children, and most



of those were comprised of families that placed a strong emphasis on English language skills. Further, several children were recruited via parent listserves for local Spanish-English immersion schools, which include children from monolingual English homes. Despite this, other studies have also included DLL groups that were English dominant (Carlson & Meltzoff, 2008).

Despite the goal of including participants that were exposed to only English, or only English and Spanish, parents reported exposure to a language other than English or Spanish for some of the children ( $N = 18$ ), including Greek, Urdu, French, Japanese, Mandarin Chinese, Farsi, Italian, Portuguese, Korean and baby sign language. Half of these children were part of the monolingual group. However, for most (including all monolinguals), exposure was minimal (hearing a few phrases once in a while from relatives; class once a week for one hour or less; hearing a few phrases at weekly karate or tae-kwon-do class) or not really a language (baby signs). Other studies have reported similar minimal exposure (and in some cases, even more exposure) to a non-native language for a monolingual group (Carlson & Meltzoff, 2008; Martin-Rhee & Bialystok, 2008). Three of these children (all in the DLL or balanced bilingual group) had more substantial exposure to a third language, including frequent Chinese and Greek classes ( $N = 1$ ), two years in an all Italian preschool ( $N = 1$ ), and exposure to French from birth ( $N = 1$ ). Two of these parents reported that the child had fair or poor understanding of the language and one parent reported the child had “ok” understanding and speaking ability in the third language.

Table 1 presents descriptive information for the whole sample, as well as broken down by language group.

Table 1. Descriptive information, overall and by language group.

	Overall <i>N</i> = 79	Bilingual <i>n</i> = 46	Monolingual <i>n</i> = 33	Balanced <i>n</i> = 17	DLL <i>n</i> = 29
<b>Background Variables</b>					
Gender, <i>N</i> (%)*					
Male	39 (49.4%)	23 (50.0%)	16 (48.5%)	4 (23.5%)	19 (65.5%)
Ethnicity, <i>N</i> (%)*					
White	43 (55.8%)	15 (34.1%)	28 (84.8%)	4 (25.0%)	11 (39.3%)
Hispanic	29 (37.7%)	28 (63.6%)	1 (3.0%)	12 (75.0%)	16 (57.1%)
Other	5 (6.5%)	1 (2.3%)	4 (12.1%)	0 (0%)	1 (3.6%)
Age (mon) <i>M</i> ( <i>SD</i> )	74.67 (9.96)	75.37 (10.33)	73.70 (9.50)	79.06 (10.90)	73.21 (9.52)
Birth Order <i>M</i> ( <i>SD</i> )	1.51 (.65)	1.51 (.67)	1.50 (.62)	1.27 (.46)	1.64 (.73)
# Children <i>M</i> ( <i>SD</i> )	2.31 (.72)	2.40 (.83)	2.19 (.54)	2.07 (.70)	2.59 (.84)
Annual Income* <sup>1</sup> <i>M</i> ( <i>SD</i> )	5.40 (1.14)	6.04 (1.76)	7.00 (1.00)	<b>5.81 (1.68)</b>	6.17 (1.81)
Mother Ed <sup>2</sup> <i>M</i> ( <i>SD</i> )	5.40 (1.14)	5.31 (1.26)	5.52 (.97)	5.44 (.96)	5.44 (.96)
Father Ed* <i>M</i> ( <i>SD</i> )	5.65 (1.20)	5.39 (1.40)	<b>6.00 (.75)</b>	5.60 (1.18)	5.28 (1.51)
Marital Stat, <i>M</i> (%)					
Married/Cohabit	71 (89.9%)	39 (84.8%)	32 (97.0%)	14 (82.4%)	25 (86.2%)
Single Parent	8 (10.1%)	7 (15.2%)	1 (3.0%)	3 (17.6%)	4 (13.8%)
<b>Language Variables</b>					
PPVT% <i>M</i> ( <i>SD</i> )	73.04 (23.35)	69.17 (23.39)	78.47 (22.54)	70.17 (20.77)	68.62 (25.06)
Eng Total Word <i>M</i> ( <i>SD</i> )	336.45 (172.75)	348.51 (167.16)	318.94 (181.89)	389.00 (230.52)	323.93 (111.31)
Eng Unique Word <i>M</i> ( <i>SD</i> )	102.82 (36.57)	104.24 (34.90)	100.74 (39.37)	111.29 (43.19)	99.96 (28.78)

\**p* < .05, (significant comparisons are bolded)

<sup>1</sup>Income: 5 = \$60,001 - \$80,000; 6 = \$80,001 - \$120,000; 7 = \$120,001 - \$175,000

<sup>2</sup>Mother/Father Ed: 5 = Bachelor's degree, 6 = Master's Degree, 7 = Doctoral Degree

For the entire sample, about half of the participants were male and average age was 74.67 months (range = 58 to 96 months). Average annual family income was high for this sample (\$60,000 – \$80,000; typical for the region in which the study took place; range = \$15,001 – \$175,000+). Participating families had 2 children on average and most participants were 1<sup>st</sup> or 2<sup>nd</sup> born (92%). One child was missing family income data, five were missing family size information, and four were missing birth order information. In terms of ethnicity, the majority of participants were White or Hispanic (including some who identified as half Hispanic and half some other ethnicity), and a few were classified as Other (including one Indian-American, two Asian, and three Biracial). Two participants were missing ethnicity information, although both of these participants were part of the bilingual/DLL group and had at least one parent born in a Latin-American country. Eleven pairs of participants were siblings: five sets within the monolingual group, three sets within the balanced group, two sets within the DLL group, and one set from both the balanced and DLL groups. Analyses were conducted a second time with one child removed from each sibling pair, and the pattern of results was the same or very similar, so siblings were retained for analyses to maintain sample size and power.

Thirty-six (46.15%) children were immigrants, defined by the child and/or at least one parent being born in a country other than the U.S. Within this group, 28 (80.0%) were second-generation immigrants (one parent born in another country, but child born in U.S.) and seven were first-generation immigrants (child born outside U.S.). One child could not be classified by generation due to missing country of birth information, and

another could not be classified as immigrant or non due to missing information for parents' country of birth. Most parents were married/remarried/cohabitating, and the remainder were either single or divorced/separated/widowed. Most mothers and fathers had a Bachelor's, Master's, or Doctoral-level degree. The remainder had either a high school diploma, some college (but no degree), or a technical degree/certification. Two fathers and one mother did not have education data. Most participants were in kindergarten (41.6%,  $N = 32$ ) or 1st grade (31.2%,  $N = 24$ ) at the time of data collection, with an additional ten (13.0%) in preschool and eleven (14.3%) in 2nd grade. Two participants were missing grade information. Eight (10.1%) children had been previously diagnosed with or evaluated for a disability, including oral motor development ( $N = 1$ ; no longer an issue for child), sensory integration/processing disorder ( $N = 2$ ), speech/fine motor delays ( $N = 2$ ), speech dyspraxia ( $N = 1$ ), and ADD/ADHD ( $N = 2$ ). The three children with some type of parent-reported speech disorder/delay were retained for analyses, as they were all monolinguals and all tested above the 50th percentile on English receptive skills.

## **Procedure**

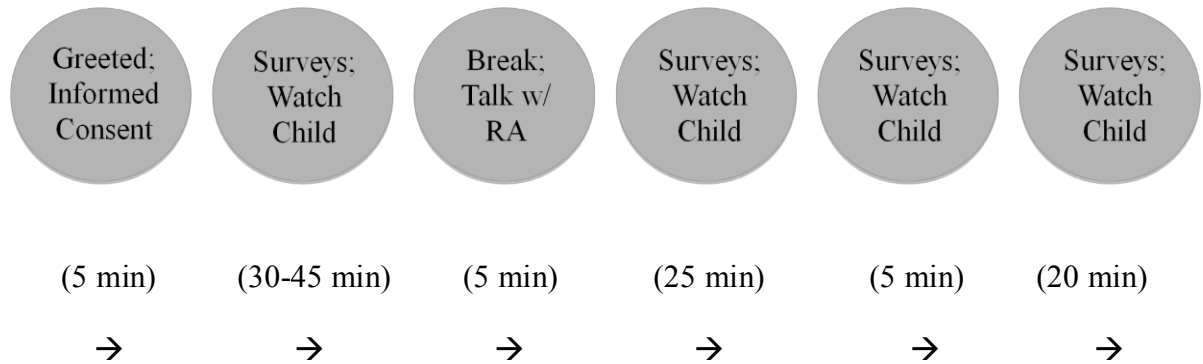
Participants were recruited through several means: 1) flyers posted at public locations, including a university campus, public libraries, and grocery stores; 2) flyers posted at a clinic that offers counseling services to families; 3) flyers distributed at several childcare centers, Hispanic/Latino churches, several psychology undergraduate classes, and to various university employee groups; 4) electronic flyers distributed to email listserves through several children's organizations and parent groups; and 5)

through word of mouth (parents of participants passed study information along to their friends). Children were compensated for participation with a small toy and/or book, along with a certificate describing them as a “junior scientist.” Parents were compensated with a \$10 gift card and a DVD copy of their child completing the tasks. The lead researcher gained IRB approval for all procedures before the study was implemented.

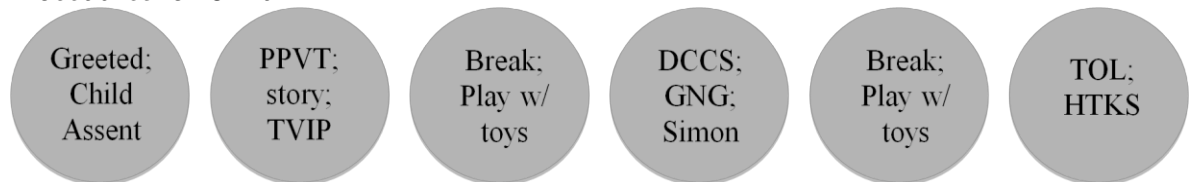
If families initiated contact via email or postal mail in response to an email, letter, flyer, or other informational posting, the researcher responded with a request that the potential participant take part in a brief phone interview with the researcher. If families initiated contact via phone, the phone interview was conducted at that time. The phone interview had four main purposes: 1) to determine eligibility (i.e. the child had to be monolingual or have exposure to both Spanish and English), 2) to learn more about the language background and experience of the target child, so that the researcher could determine in advance what language group the child would likely fall into (bilingual/DLL or monolingual), 3) to determine what language was most comfortable for the child according to parents so that EF assessments could be conducted in that language (all in Spanish or all in English) and bilingual research assistants could be scheduled as needed, and 4) to schedule a date and time for the child and parent to come to the university research lab for the study. A graduate research assistant fluent in both Spanish and English was available to conduct the phone interview in Spanish, as needed. Five families had the phone interview conducted in Spanish, one of whom was excluded due to children’s low language scores (described above).

On the day of participation, the child and parent were greeted outside by the RA(s) and led into the assessment room. Figure 1 depicts the order of all procedures for the parent and child.

*Procedures for Parent*



*Procedures for Child*



Note: Surveys = Language Background Questionnaire, Behavior Rating Inventory of Executive Function, Child Behavior Checklist; PPVT = Peabody Picture Vocabulary Test; Story = Story telling task (one for monolingual, two for bilingual); TVIP = Test de Vocabulario en Imágenes Peabody; DCCS = Dimensional Change Card Sort; Simon = Simon task; GNG = go/no-go task; TOL = Tower of London; HTKS = Heads-toes-knees-shoulders task

Figure 1. Order of experimental procedures for parent and child.

Table 2 describes all measures and what each was assessing. For participants expected to be DLL/bilingual (based on the phone interview), a Spanish-English bilingual RA (either a psychology graduate student or undergraduate honors student) was present to conduct the session. On the day of testing, the RA inquired again as to the language that the parent is most comfortable with, and this language was used with the parent throughout

the study. The parent was given the option of completing all surveys in either Spanish or English. Nineteen parents chose to complete the surveys in Spanish, and the remainder completed them in English.

Table 2. Description of all measures of EF, behavior, and language.

Assessment Name	Abbreviation	Purpose of Assessment
<b>Language Assessments</b>		
Peabody Picture Vocabulary Test (PPVT)	PPVT	English receptive language skills
Test de Vocabulario en Imágenes Peabody	TVIP	Spanish receptive language skills
Narrative Elicitation Task	N/A	Spanish and/or English productive language skills
Language Background Questionnaire	N/A	Background of language experience (parent report)
<b>Executive Function Assessments</b>		
Dimensional Change Card Sort	DCCS	Cognitive flexibility (set shifting)
Simon Task	N/A	Interference control
Go/No-Go Task	GNG	Inhibition of a prepotent response
Tower of London	TOL	Planning skills
Head-Toes-Knees-Shoulders Task	HTKS	Inhibition of a prepotent response
Behavior Rating Inventory of Executive Function	BRIEF	Overall executive function (parent report)
<b>Behavior Assessment</b>		
Child Behavior Checklist	CBCL	Behavior problems (parent report)

After being seated in the assessment room, the parent was given the informed consent to review and sign while the child played with available toys. The RA did rapport-building with the child at this time (as well as during the walk from the parking location). Then the

parent was taken to a separate room where they could watch/hear the child on a television monitor. At that time, the RA gave the parents three surveys to complete at their leisure during the session. The RA then returned to the assessment room with the child and continued to converse/play with the child for a few minutes to ensure that he/she felt comfortable. For bilingual children, the RA switched back and forth between Spanish and English to gather anecdotal evidence regarding the child's comfort with speaking each language. Then the RA had the child sit with him/her at a small table and began the study by gaining child assent.

After gaining child assent, three (monolingual) or four (bilingual) language assessments were administered, including measures of receptive English skills (Peabody Picture Vocabulary Test; PPVT; Dunn & Dunn, 1997), receptive Spanish skills (Test de Vocabulario en Imágenes Peabody; TVIP; Dunn, Padilla, Lugo, & Dunn, 1986), and expressive English and Spanish skills (a story telling task; done once in English and for bilinguals only, once in Spanish). Monolinguals, in addition to bilinguals, were administered the Spanish vocabulary test to ensure that they did in fact have limited knowledge of Spanish. The order of the language assessments was always a vocabulary test (PPVT or TVIP) followed by the story telling task, followed by the second story telling task (bilinguals only), and finally the second vocabulary test. The order in which each language was assessed varied across bilingual participants. Twenty-eight received the English assessments first, and the other 18 received Spanish first; Spanish was only administered first to children thought to be fairly comfortable with this language. For monolinguals, the English vocabulary test was always administered first. In addition, two



books were used for the story telling task (described in detail below). The presentation of the books in terms of order (bilinguals) or which book was used (monolinguals) also varied across participants. Twenty-three bilinguals had *Frog Goes to Dinner* first (eight of which did the story in Spanish) and the other twenty-three had *Frog Where Are You* first (10 of which did the story in Spanish). For monolinguals, 16 told their story with *Frog Goes to Dinner* and the other 17 told their story with *Frog Where Are You*. These tasks lasted approximately 30 minutes for monolinguals and 45 minutes for bilinguals.

For balanced bilinguals only, there was no main effect for the type of book used,  $F(1, 13) = 2.64, p = .128$ , or for the order of language presentation,  $F(1, 13) = .01, p = .128$  on unique English words produced. There was also not an interaction between these two variables,  $F(1, 13) = .57, p = .463$ . Findings followed the same pattern for total English words, unique Spanish words, and total Spanish words. In other words, the type of book used (*Frog Goes to Dinner* or *Frog Where Are You?*) and the order of language assessment (either English first or Spanish first) did not appear to influence word production. However, balanced bilinguals did seem to produce significantly more words in English ( $M = 341.75, SE = 32.02$ ) compared to Spanish ( $M = 283.25, SE = 27.48$ ), no matter the order of language assessment,  $F(1, 14) = 13.73, p < .05$ . (DLLs could not be included in these analyses, as most of them said little to no words in Spanish, which was expected given their dominance in English). When looking at all children, the total number of English words produced was not influenced by the book the child used for the English task,  $F(1, 72) = .002, p = .965$ , nor were unique words,  $F(1, 74) = 1.67, p = .20$ .

Following the language assessments, the child was given a five-minute play break. Toys were available in the assessment room for children to play with. During this first break, the RA determined which language appeared to be dominant for the child based on several factors: 1) answers given by the parent during the phone interview, 2) the RA's anecdotal observations of child language use during the initial rapport-building session and language assessments, and 3) the child's response to the RA's direct questions about language of preference. If the RA's anecdotal observations agreed with parent report, all EF assessments were given in that language. If the parent indicated that the child was equally strong in both languages, and the RA agreed, then the RA defaulted to English for all EF assessments. If the RA's anecdotal observations of language did not match parent report (though this never occurred), the RA was instructed to use the preference indicated by the child at the time of the study.

Next, the child completed three EF assessments. First was a measure of cognitive flexibility – the Dimensional Change Card Sort – followed by a measure of interference control (the Simon task), and then a measure of inhibition of a prepotent response (go/no-go task). Then the child took another five-minute play break. Finally, the child completed three additional EF tasks: a planning task (the Tower of London), another measure of inhibition (Head-Toes-Knees-Shoulders task) and a task to elicit private speech (a selective attention task). The selective attention task was included in the current study so that private speech for this unique sample could be analyzed in future studies; results from this task were not considered in the current dissertation. The EF assessments took about 45 minutes to complete. In between tasks, the RA went to speak briefly with the

parent to see if he/she had any questions about the surveys. At the end of the tasks, the child was reunited with the parent and then was able to pick a prize for participating. The parent was given the gift card at this time. The DVD was mailed at a later date. In total, the study lasted approximately 1.5 – 2 hours. All child assessments were video-recorded with a digital camera that sat on a tripod in the room with the child. The researcher explained that the camera was making a movie so that the child could watch him/herself later. The camera was moved around the room four times between assessments to capture child task performance from appropriate angles. All children appeared comfortable with the presence of the camera and most ignored it after the tasks began.

### **Training of Research Assistants**

All research assistants (RAs) attended training sessions with the author. RAs included both bilingual and monolingual graduate students and undergraduate students, all in psychology. Across several training sessions, most of the RAs were trained to administer all of the measures, including the PPVT, TVIP, story-telling task, and EF measures. A few others were trained only to administer the language assessments (due to high demands on bilingual RAs' time). They were provided with scripts to use for each assessment and they practiced administering the assessments with one another, and then with pilot children. RAs were blind to specific hypotheses of the study and also did not know prior to assessments if the child had externalizing behavior problems. However, it was not possible to prevent bilingual RAs from knowing the dominant language (English vs. Spanish) of bilingual participants, as this was required knowledge for administering the assessments appropriately.

## Measures

### Receptive Language Skills

The third edition of the Peabody Picture Vocabulary Test (PPVT-III; Dunn & Dunn, 1997) was used to directly assess children's receptive language skills in English. The PPVT can be administered to individuals ranging in age from 2½ years through adulthood. It is available in two forms, Form IIIA and Form IIIB. Children's results did not vary based on the PPVT form that was used, for both standard scores,  $t(75) = -1.28, p = .21$ , and percentile ranks,  $t(75) = -1.35, p = .18$ . Form was counter-balanced across participants; 41 received Form A and 38 received Form B. Each form contains four training items and 204 total possible test items (17 sets of 12 items each). The items increase with difficulty throughout the test. Each item or "picture plate" has four pictures on it. For each plate, the test administrator asks the participant to "point to" or "show me" a target word. Testing time averages around 12 minutes in length. The test is complete when a participant makes eight or more errors within a set of 12 items (ceiling set). The participant's basal is the lowest set of items containing one or no errors. Raw scores are tallied at the end of test administration by subtracting the total number of incorrect items from the ceiling item. Raw scores are converted into standard scores, percentiles, and age equivalent scores by using norms tables. Split-half reliability coefficients for the PPVT-III, based on a standardization sample of 2,275 individuals (most of whom were children under age 19) were very high, ranging from .86 to .97 (Kaufman & Lichtenberger, 2006). The PPVT-III also correlates highly with other measures of intelligence, indicating good validity.

All participants, even monolinguals, were administered the Test de Vocabulario en Imágenes Peabody (TVIP; Dunn et al., 1986), the Spanish version of the PPVT-Revised (Dunn & Dunn, 1981). The TVIP is administered in the same manner as the PPVT, where a child is shown picture plates with four items on each plate. The researcher asks the child to point to a target word, and the child must indicate the correct picture out of the four choices. However, establishing a basal and ceiling is different from that for the PPVT. For the TVIP, the basal is the highest eight responses in a row containing no errors. The ceiling is lowest eight responses in a row containing six or more errors. Raw scores are calculated by subtracting the number of errors from the ceiling item. Raw scores are then converted to standard scores, percentiles, and age equivalent scores using norms tables. Though monolinguals were expected to have extremely limited or no knowledge of Spanish, they also received the TVIP to ensure that they were in fact monolinguals. As expected, several monolingual children reported to the RA that they did not know any Spanish. The RA reassured them that it was just a game to see how many Spanish words they might know, that it was ok to guess, and that the game would go quickly. After these reassurances, all monolingual children appeared comfortable completing the Spanish vocabulary assessment. The TVIP was normed on samples of Spanish-speaking monolingual children in Mexico and Peru. The median split-half reliability coefficient was .93. Percentiles from both the PPVT and TVIP were used to help calculate children's scores for "degree of bilingualism" (see section on Calculating Degree of Bilingualism).

## **Expressive Language Skills**

Children's expressive language skills were assessed through an open-ended storytelling task, where children are asked to look at a wordless picture book and produce an original narrative. Wordless pictures (in the form of picture books or sequences of picture plates) have been used frequently in prior studies to elicit narratives from young children (Cain, 2003; Cain & Oakhill, 1996; de Villiers, 2004; Shapiro & Hudson, 1991). The current study included this narrative elicitation task in addition to standardized measures of vocabulary (PPVT; TVIP) because narrative samples are high in ecological validity and they mirror the ways in which children would engage in language in everyday settings (Bedore, Peña, Gillam, & Ho, 2010; Westby, Van Dongen, & Maggart, 1989).

For monolingual participants, the task involved one wordless picture book, either "Frog, Where Are You?" (Mayer, 1969) or "Frog Goes to Dinner" (Mayer, 1974). These same books have been used in prior studies to elicit narratives from children ranging in age from 3 to 10 years (Bedore et al., 2010; Botting, 2002; Greenhalgh & Strong, 2001; Heilmann, Iglesias, & Nockerts, 2008; Justice, Bowles, Pence, & Gosse, 2010; Miller et al., 2006; Muñoz, Gillam, Peña, & Gulley-Faehnle, 2003; Norbury & Bishop, 2003; Simon-Cerejido & Gutiérrez-Clellen, 2009). According to Justice and colleagues, these books are appropriate for collecting children's narratives, as the pictures provide an obvious plot structure. For DLL/bilingual participants, the task involved both books in order to elicit two different narratives, one in English and one in Spanish. Several prior studies have looked at children's narratives based on a picture book with a sample of Spanish-English bilingual children (Bedore et al., 2010; Muñoz et al., 2003; Miller et al.,

2006; Simon-Cereijido & Gutiérrez-Clellen, 2009; Uccelli & Páez, 2007). Further, prior research has indicated that narrative samples taken in two languages from DLLs/bilinguals can be accurately and reliably transcribed so that comparisons can be made across languages (Heilmann et al., 2008).

According to Miller and colleagues (2006), it is preferable to have children tell their narratives in each language using the *same* picture book, so that the narratives can be compared on several different features across stories. However, Miller et al. had a delay of one week between the elicitation of children's Spanish and English narratives to preserve the independence of the languages during the task. In the current study, there was not a time delay between the two narratives; the DLL/bilingual children relayed their narratives one right after the other. To deal with this issue, I chose to use two different Frog books. Other researchers have also used multiple Frog books when working with bilingual participants (Bedore et al., 2010; Simon-Cereijido & Gutiérrez-Clellen, 2009) or when eliciting multiple narratives from monolingual participants (Greenhalgh & Strong, 2001) to good effect.

The procedure for the story-telling task followed the Narrative Elicitation Protocol outlined by Justice and colleagues (2010). This elicitation protocol is part of a larger Narrative Assessment Protocol described by Justice et al., but which was not used in the current study. To begin, the RA told the child that the RA would go through the pictures in the book once with the child, and then it would be the child's turn to create a make-believe story about the book. The RA then went through the book once with the child, pointing out the pictures, and spending eight to ten seconds on each page. The RA

did not make any comments about the content of any of the pictures. If the participant made a comment, the RA reminded the child that they were only looking at the pictures at that time. After going through the book one time, the RA handed the child the book and asked him/her to tell a make-believe story using the pictures in the book. The child was told to make the story as long and interesting as possible (Justice et al., 2010). While the child told his/her story, the RA only provided positive feedback (“Good job!” or “I like your story!”), repeated exactly what the child had just said, or prompted the child to continue if the child appeared stuck on a page. When the child reached the end, the RA asked if the child had anything else to add. For bilingual participants, the RA repeated the task, including all instructions and feedback, in the opposite language.

### **Transcribing and Coding of Elicited Narratives**

All English narratives were transcribed word-for-word into Word documents by an RA and were then verified by the lead researcher. Bilingual RAs transcribed and verified the Spanish narratives. All RAs were trained in the transcription process by transcribing five of the same stories, and then meeting to discuss any discrepancies in their transcripts and making final decisions about what the child said. Transcriptions were coded by bilingual (Spanish transcripts) and monolingual (English transcripts) RAs after verification was complete. RAs were trained on the coding directions, and then two RAs each coded the same five transcripts and compared results. Any discrepancies were discussed and final decisions were made regarding the codes. For English story coding, inter-rater reliability was extremely high for both total word count ( $r = .997$ ) and unique/different word count ( $r = .969$ ; both of these counts explained more below). For



Spanish story coding, inter-rater reliability was also very high for both total word count ( $r = 1.00$ ) and unique/different word count ( $r = .984$ ). The rest of the transcripts were coded by one RA each.

Transcriptions were coded for two different features: total number of words uttered by the child and number of different words (NDW). For total number of words, any and all intelligible words the child uttered were part of this code category, including grammatical errors, repeated words and restarts, but excluding filler words like “um” or “uh” or words/phrases that were clearly directed to the self or to the experimenter and were not part of the story (ex: child says “What is that?” to self or experimenter about a picture in the book). For NDW, the RA made a list of the unique words the child uttered across the entire transcript (Bedore et al., 2010). For instance, once a child said “frog” in his/her story, the word “frog” would never count again toward the NDW total. NDW is thought to reflect a child’s language growth (Bedore et al., 2010). For instance, compared to their typically developing peers, children diagnosed with language impairment tend to have lower NDW scores (Hewitt, Hammer, Yont, & Tomblin, 2005; Redmond, 2004). NDW has been used frequently when coding narratives in prior research (Muñoz et al., 2003; Miller et al., 2006; Uccelli & Páez, 2007; Simon-Cerejido & Gutiérrez-Clellen, 2009). See the section on Calculating Degree of Bilingualism for further details on how these scores were used.

### **Language Experience and Exposure**

The Language Background Questionnaire created by Ellen Bialystok (Bialystok, 2010; Bialystok & Feng, 2009; Bialystok & Viswanathan, 2009) was administered to

parents, with some modifications from the original version, to gather information regarding children's language experience and exposure, as well as child and family demographics. Several items from the survey are reported as descriptives about the sample, and other items were used to help determine if a given child was a balanced bilingual or a DLL (see below).

For the current study, the survey was four pages long, and was available in both English and Spanish, so that parents could complete it in their language of preference. For Part A, parents answered questions about child and family demographics. Bialystok's original survey asked for the following demographic information: child date of birth, child gender, child country of birth, child handedness, date when child first came to U.S., country of birth for each parent, date when each parent came to U.S., education level for each parent, and occupation for each parent. For the current study, several more demographic questions were added, including child ethnicity, number of children in the family, birth order of target child, marital status of parents, annual family income, whether or not child has a disability diagnosis, type of diagnosis, whether or not child has an IEP, what special need IEP was created for, and an additional option under education level called "Technical Degree or Certification."

For Part B of the survey, parents answered questions about the languages the child speaks, when and where the child learned the languages, where the child uses the languages, and how proficient the child is at understanding and speaking the languages. Features that I added or modified from the original survey include: two questions regarding the child's ability to speak and understand English (original survey only asked

about languages other than English), two additional response options for the proficiency questions (“Good” and “OK” options; original survey had only “Excellent,” “Fair,” and “Poor”), and two questions asking the age at which the child first began to learn each language. These modifications were made so that I could gather more information to help classify children as balanced bilinguals, DLLs, or monolinguals (see below). Parents’ responses to children’s understanding and speaking ability in one (monolinguals) or both (bilinguals) languages was converted to numbers, where Excellent = 5, Good = 4, OK = 3, Fair = 2, and Poor = 1. A proficiency score for each language ranging from 1 (poor) to 5 (excellent) was calculated by averaging together the scores from the two statements regarding receptive and expressive ability, resulting in a parent-reported English proficiency score and a Spanish proficiency score (bilinguals only).

Part C of the survey asked parents about the child’s exposure to the two languages in terms of input and output. Each statement required a response ranging on a 1 to 5 scale, with 1 = “Always English,” 2 = “Mostly English,” 3 = “Both Languages Equally,” 4 = “Mostly Spanish,” and 5 = “Always Spanish.” This scale reflects actual language use by the child in his/her every day environment. For the current study, I added or modified the following statements: changed statement from “language spoken by child to parents” to say “language spoken by child to mother;” added two additional statements about the language spoken by child to father and to other adult relatives; modified statement “language spoken by you to your child” to say “language spoken by mother to child;” added two additional statements about the language spoken by father and other adult relatives to child; added two statements about the language spoken by child in school and

language spoken by teacher in school; and added a statement about the language in which child reads stories.

Seven of these statements reflect language *output* for the child and seven reflect language *input* for the child. The last two statements were part of Bialystok's original survey, and were more for descriptive purposes, as they do not clearly reflect language input or output for the child (asked about language in which parents watches TV and reads books). Children's scores on the first fourteen statements (reflecting the balance between the two languages for input and output) were averaged together to get one score for children's language use/balance. Additionally, based on recommendations from prior research that it is important to focus on bilingual children's actual language use (Morton & Harper, 2007), a more stringent language use score was calculated based on only six of the seven language *output* items (excluding one item that asked about the language in which the child reads). So, because this scale ranged from 1 to 5, with 1 = All English and 5 = All Spanish, a score in the range of 2.5 to 3.5 indicated almost complete balance between the two languages in terms of actual use. A score less than 2.5 would indicate greater English language use by the child, and a score above 3.5 would indicate greater Spanish language use by the child.

### **Determining Child Language Group**

Six criteria were used to determine language group, including both parent report and scores from direct assessments. The child had to meet five of six criteria to be considered a balanced bilingual. If they did not meet two or more of the criteria, the child was considered a DLL, provided that they had at least 6 months of substantial exposure

(e.g. exposed to and/or using L2 several times per week) to the second language. To be considered a monolingual, parents reported that the child did not speak or understand any language other than English (excluding minor exposure, such as a 30 minute class one time per week). The six criteria to be considered a *balanced bilingual* were as follows:

1. Minimum exposure of three years (or more) to *each* language.
2. Average parent-reported proficiency scores of 3 or higher for *each* language.
3. Average parent-reported score of 2.5 to 3.5 for language use, indicating almost equal use of Spanish and English (for either the original use score or the more stringent use score).
4. PPVT age equivalency score within the instrument's confidence interval window for the child's actual age in months (or higher).
5. TVIP age equivalency score within the instrument's confidence interval window for the child's actual age in months (or higher).
6. English and Spanish unique word counts in the story differ by 50 words or less.

The above criteria for language group were created based on prior research. First, when thinking about *how* bilingual an individual is, researchers have noted the importance of length of exposure, proficiency or ability in each language, and actual use of the two languages (Baker 2006; Bialystok, 2001). Criterion 1 taps into length of exposure, criteria 2, 4, 5, and 6 tap into language proficiency/ability, and criterion 3 taps into actual language usage. Second, the minimum requirement of three years of exposure has been used as a criterion in other studies (Guitérrez-Clellen & Kreiter, 2003), and is based on findings from Cummins (1984) that three to four years of exposure to a

language is required “to achieve basic social language competencies” (Guitérrez-Clellen & Kreiter, 2003, p. 271). Third, the computation of the balance between the two languages by averaging scores across the parent-reported statements of language use is based on prior work by Bialystok (Bialystok, 2010; Bialystok & Feng, 2009).

In the current study, as described above, I averaged scores from the 14 statements about language use (ranging from *All English* to *All Spanish*), to get one score for language use. I used more lenient cut-offs for what is considered a “balanced bilingual” (see criterion 3 above). In prior studies, Bialystok and colleagues (Bialystok, 2010; Bialystok & Feng, 2009) considered a score of 3 to be fully balanced, and deviations from this to indicate a bias for one language or the other. However, as already noted, language exposure (input) and production (output) are only part of the broader picture of language experience for a bilingual child; the more lenient cut-offs in the current study (see criterion 3) allow a bilingual child who has a score slightly above or below a 3 to still be considered balanced, provided that the other criteria are met.

### **Calculating Degree of Bilingualism**

“Degree of bilingualism” was calculated to provide a continuous measure of *how* bilingual each child was (excluding monolinguals). To my knowledge, this has never been done in prior empirical research. Several steps were taken to create this variable. First, three difference scores were calculated: one for parent report of English and Spanish proficiency, one for percentiles scores from the PPVT and TVIP, and one for the NDW counts from the English and Spanish stories. The absolute value was calculated for each of these three variables so that bigger numbers indicated less balance (a greater

difference) in English and Spanish receptive and expressive language skills. Then a similar score (ranging from 1 to 3) was created based on the language use variable: middle scores of 2.5 to 3.5 were turned into 1's, more distant scores of 1.5 to 2.4 or 3.6 to 4.5 were turned into 2's, and extreme scores of 1 to 1.4 or 4.6 to 5 were turned into 3's. This means that a score of 3 indicated less balance in terms of language use and a score of 1 indicated about equal use of Spanish and English by the child. These four new variables (parent-reported language proficiency difference score, receptive language difference score, expressive language difference score, and recoded language use score) were standardized by creating z-scores for each, and were then all averaged together to create the final variable "degree of bilingualism." For this final variable, the larger the score for a given child, the *less* balanced that child was across his/her two languages.

### **Executive Functioning: Cognitive Flexibility**

The Dimensional Change Card Sort (DCCS; Frye et al., 1995) is often used to assess cognitive flexibility and set shifting, and is "perhaps the simplest possible test of task switching" (Diamond, 2002, p. 475). The original DCCS involved first sorting cards by one sorting rule (color), and then sorting by a new rule (shape), when instructed by the experimenter. The DCCS has been used with bilingual samples (Bialystok, 1999; Bialystok & Martin, 2004), as well as children with ADHD symptoms (Fahie & Symons, 2003). This task has convergent validity, as demonstrated by moderate correlations with academic skills ( $r = .42$  to  $.63$ ) and teacher ratings of child behavior ( $r = .56$ ; DCCS performance was one of several tasks that loaded highly onto the same factor; Lipsey, Wilson, & Farran, 2010). It also correlated well with later academic skills ( $r = .43$  to  $.64$ ),

indicating good predictive validity. However, many 4- and 5-year-olds perform at ceiling on the DCCS, so a more advanced “star” version was developed for use with children up to 7 years of age, and this version was used in the current study (Carlson, 2005; Zelazo, 2006). Prior research has indicated that this advanced version is difficult for older children to complete successfully (Carlson, 2005).

Children completed the task manually. Procedures primarily follow those outlined by Zelazo (2006), with a few additions from Carlson and Meltzoff (2008) and Zelazo (2006). For the current study, I followed the original number of block 3 trials recommended by Zelazo (2006), the developer of the advanced DCCS. However, instead of block 3 cards having a black border around them, they had a star on the front of the card (following Carlson & Meltzoff, 2008); this was thought to be a more noticeable dimension for children.

At the start of the task, the RA presented the child with two small boxes. One box had a red boat on a wooden panel attached to the box with Velcro (visible to the child), and the other box had a blue rabbit attached. The child was told that first they were going to play the shape game. Then the RA told the child the instructions for sorting the cards by shape: “All rabbit cards go here” (RA pointed to box with blue rabbit on it) “and all boat cards go here” (RA pointed to box with red boat on it). For all target cards, the color was opposite that of the cards attached to the boxes (all target cards were red rabbits or blue boats). The RA sorted one card (a boat) as an example and asked the child to sort another card (rabbit) as a second example. Then the RA went through the first block of test trials (the “preswitch” block), handing the child six target cards to be sorted on shape.



The child was not given the same type of card to sort more than twice in a row (following Zelazo, 2006). On each trial, the researcher labeled the card based on dimension (“Here is a rabbit”), handed the card to the child, and asked the child where to put it. Feedback on correctness was not provided during test trials. The sorting rule for block 1 (“All rabbits go here and all boats go here”) was repeated twice, once before test card 1 and again before card 4. [It should be noted that the statement of the sorting rule every three cards, in both this first block and the next two, does not follow what other researchers have done. Other researchers repeat the rule every trial; however, I felt this would make the task too easy, especially since I was testing children over 7 years of age].

Then the RA told the child that they were done with the shape game, and would now play the color game. The researcher explained that the new rule is to sort the cards by color – “All red cards go here” (pointing to box with red boat) “and all blue cards go here” (pointing to box with blue rabbit). Then block 2 (the “post-switch” block) proceeded in the same manner as block 1, but with no practice. This time, the RA labeled each target card based on color (“here is a red one”). The RA went through six test trials, handing the child the cards one at a time to be sorted.

Lastly, the RA moved onto the third and final block of the task. This block was unique to the Advanced DCCS. It involved using both sorting rules during the same block of trials. The relevant dimension to sort on for a given card was indicated by the presence or absence of a star on the card (Carlson, 2005; Carlson & Meltzoff, 2008). At the start of this block, children were told that they were done with the color game and would now be playing the star game. The RA explained that if a card had a star on it, the

child should sort the card by color, and the RA then proceeded to sort a star card as an example. Then the RA explained that if there is *not* a star on the card, the child should sort the card by shape, and sorted a card without a star as an example. The RA then did a rule check to make sure the child understood when to sort by color and shape. If the child did not give the correct answers about the sorting rules when asked, the RA repeated the rule and questioned the child again about the rule. Then the RA went through twelve test trials with the child in the same manner as the first two blocks. On each trial, the RA labeled the card (“here is one with a star”), handed it to the child, and asked him/her where to put it. The sorting rule (“If there is a star, play the color game, but if there is no star, play the shape game”) was repeated four times: before test cards 1, 4, 7, and 10. Six cards had a star and six did not. The cards were presented in the same order for all children. Following Carlson and Meltzoff, the score from this task was total number of star errors made during Block 3 (ranging from 0 to 6).

### **Executive Functioning: Interference Control**

The Simon task was used to assess interference control – a child’s ability to deal with conflicting perceptual information. This task has been used with bilingual samples (Bialystok et al., 2004; Martin-Rhee & Bialystok, 2008), as well as children with ADHD symptoms (see Mullane, Corkum, Klein, & McLaughlin, 2009 for a review). The task has been used with children ages 4 and up.

For the current study, procedures follow Martin-Rhee and Bialystok (2008), with a few changes noted. The task was programmed and completed on a 15-inch laptop computer using E-prime software. It included two stimuli – a red square and blue square.

Prior to the start of the task, a red sticker was affixed to the “A” key on the keyboard and a blue sticker was affixed to the “L” key. Participants were told to push the red sticker on the left of the keyboard whenever a red square appeared on the screen, and to push blue sticker on the right whenever a blue square appeared on the screen. Children were told to respond as quickly and accurately as possible. For some trials, the colored square was presented on the same side of the computer screen as the correct button to push (*congruent trials*). These trials only required the child to attend to the *color* of the stimulus. For other trials, the colored square appeared on the opposite side of the screen from the button to be pushed (*incongruent trials*). These trials had perceptual conflict, as the child would attend to both the type of stimulus and stimulus location. Participants typically respond slower on incongruent trials compared to congruent trials – the ‘Simon effect.’ This effect is both robust and reliable (Simon, 1990), though some have found that it is only reliable when the proportion of congruent to incongruent trials is high (discussed further below; Borgmann, Risko, Stolz, & Besner, 2007).

Following Martin-Rhee and Bialystok (2008), after receiving instructions, the child began the task with eight practice trials containing a mix of congruent and incongruent trials. The child had to make no more than two errors on the practice trials in order to move on to the test trials. If more than two errors were made, the child repeated the practice set. Martin-Rhee and Bialystok did not describe feedback during the practice; however, for the current study, the child received feedback on the screen about correctness of response after each practice trial. For correct responses, a smiley face

appeared with the word “Correct!” below it. For incorrect responses, an “X” appeared with the words “Not correct” below it.

There were 60 experimental trials [Martin-Rhee & Bialystok (2008) only had 40, but their participants were 4 years old]. Fifty percent of trials were congruent in Martin-Rhee and Bialystok’s study; however, Borgmann and colleagues (2007) found that the Simon effect was more reliable (i.e. interference control more likely to occur) if the proportion of congruent trials was closer to 75%. Further, van Mourik and colleagues (2009) used a 50-50 split for the trial type with a sample of children diagnosed with ADHD. According to van Mourik et al., having a higher proportion of congruent trials (such as 75%) would make the task too difficult for children with ADHD-like symptoms due to the high strain on working memory. These authors also mentioned that pilot testing found the task to be too difficult for children under age 8, though Martin-Rhee and Bialystok used 4-year-olds. Therefore, for the current study, a compromise was made in the middle at 60% (36 congruent trials and 24 incongruent).

Following Martin-Rhee and Bialystok (2008), at the beginning of each trial, a fixation cross appeared in the middle of the screen for 500 ms to orient the child away from either the left or right side. Then the stimulus appeared on screen for 5,000 ms, or until the child responded. If no response was made, it was considered an error and the next fixation cross appeared. Feedback was not provided during experimental trials. Scores from this task included reaction time and percent accuracy for each trial type (congruent and incongruent). Lower scores indicate faster and/or more accurate performance. I also calculated the difference in reaction time (RT) between incongruent

and congruent trials (the “Simon Effect”), where smaller numbers indicate better interference control. Average RT on each trial type (congruent and incongruent) and the Simon effect score were used in data analyses. Scores from this task have demonstrated intra-individual reliability across blocks of trials ( $r = .56 - .65$ ) when the proportion of congruent to incongruent trials is .50 or .75 (Borgmann et al., 2007). The task is valid, as indicated by findings from the Simon task following the same pattern as other tasks tapping into interference control (i.e. Stroop task; van Mourik et al., 2009).

### **Executive Functioning: Inhibition of a Prepotent Response**

Inhibition of a prepotent response (simple response suppression) was assessed the go/no-go (GNG) task. GNG tasks have been used often in studies with hard-to-manage children or those diagnosed with ADHD (Barkley, 1997; Brophy et al., 2002; Iaboni et al., 1995; Shue & Douglas, 1992; Trommer et al., 1989), as well as with bilinguals (Emmorey et al., 2008). The task was conducted on the same computer as the Simon task. The RA began by telling the child that he/she should push the spacebar as fast as possible when squares with straight or diagonal lines appeared on the screen, but not when a square with an X appeared. Following Berlin and Bohlin (2002) and Brocki and Bohlin (2004), there were three “go” stimuli and one “no-go” stimulus. The stimuli were perceptually similar to increase difficulty and the need for attention to the task. Go stimuli included a square with a diagonal line to the right, a square with a diagonal line to the left, and a square with a vertical line down the middle. The no-go stimulus was a square with an X in it. Choice of stimuli followed Brocki and Bohlin, 2004, except that in their study, the square with the vertical line was the “no-go” stimulus.

The researcher showed the stimuli to the child and described them as either requiring or not requiring a response. Following Kuntsi, Andreou, Ma, Borger, and van der Meere (2005), the child then experienced five practice trials in random order, including one no-go trial. Children had to get 100% accuracy to move on to test trials. Feedback on correctness was provided during the practice (same as Simon task). Actual testing consisted of 60 trials presented randomly. Prior studies have used trial lengths ranging from 460 ms for children over age 6 (Brocki & Bohlin, 2004) to about 2,500 ms for preschoolers (Berlin & Bohlin, 2002). Therefore, I chose to follow Raaijmakers et al. (2008) with a trial length of 1,500 ms. At the start of each trial, the stimulus appeared on the screen, and then disappeared either when the child responded or after 1,500 ms. Prior research has also used a variety of inter-stimulus intervals (ISI) with the GNG. However, Shue and Douglas (1992) have pointed out the need for shorter ISIs with this task in order to place less demands on attentional resources, and more demands on inhibitory control. Therefore, again following Raaijmakers et al., I originally chose an ISI of 1,500 ms. However, pilot testing revealed that the task appeared to be too easy with such a long ISI, so it was reduced to 1,000 ms for test subjects. Following Berlin and Bohlin (2002), 70% of all trials were “go” trials, in order to build a prepotent response (Eigsti et al., 2006). Scores from this task included errors of commission (making a response to the no-go stimulus). Errors of commission are considered to be a direct measure of behavioral inhibition (Berlin & Bohlin, 2002). This task has good test-retest reliability, with inter-class correlations ranging from .5 to .9 (Kuntsi et al., 2005).

## **Executive Functioning: Planning Skills**

This aspect of EF is most often assessed with the Tower of London (TOL; Shallice, 1982) or Tower of Hanoi (TOH; Simon, 1975). The TOL was used in the current study, as it has been used more often with younger ages. This task involved a set of different colored balls (green, red, and blue) on three wooden pegs. The three pegs were different lengths, so that the first peg could hold all three balls, the second peg could hold two balls, and the third peg could only hold one ball. The goal of the task was for the participant to move the balls around one at a time so that the participant's ball-and-peg set ("starting state") matched the experimenter's ball-and-peg set ("ending state"). The participant was instructed to complete the task in the fewest moves possible. TOL problems differ in their difficulty level depending on the minimum number of moves required, ranging from two moves up to five. This type of task has been used with children with ADHD symptoms (Kopecky et al., 2005), but not bilingual children. It is good at detecting EF deficits (Klorman et al., 1999). It has been used often with children four years and older (Fernyhough & Fradley, 2005; Levin et al., 1991; Lidstone et al., 2010; Luciana & Nelson, 1998).

For the current study, procedures mostly followed Fernyhough and Fradley (2005), with a few noted exceptions. The RA began the task by placing a wooden peg set with two colored balls (blue and green) on it in front of him/herself and a matching pet set in front of the child. For each test trial, the child's peg set always displayed the "starting state" (what the child began with) and the RA's peg set always displayed the "ending state" (what the child was trying to accomplish). The RA explained to the child

that, with this game, the child wanted to make his/her set look just like the RA's set. The RA then explained three rules to the child: 1) only one ball at a time could be moved, 2) balls must always be moved to a peg, and not somewhere else, like the table, and 3) the puzzle should be solved in the fewest number of moves possible. For each rule, the RA demonstrated the correct thing to do. Following Fernyhough and Fradley, children were asked to put one hand in his/her lap during the game to minimize rule breaks, and were also encouraged to talk out loud during the game to elicit private speech (private speech was not considered in the current dissertation; it will be analyzed in future studies). After explanation of the rules, the child went through two practice trials involving only the blue and green balls. During practice, if the child broke a rule, the researcher pointed it out, reminded the child of the rules, and made the child correct the error (Fernyhough & Fradley, 2005, do not mention doing this). After the practice was complete, the RA added a red ball to each peg set and test trials began.

At the start of each trial, the RA set his/her ball-and-peg set to the ending state, and set the child's ball-and-peg set to the starting state. While setting up the problem, the RA reminded the child of the ultimate goal ("to make your peg set look just like mine") and also encouraged the child to *think* about the problem before he/she actually started moving the balls (Levin et al., 1991; Luciana & Nelson, 1998), a slight deviation from Fernyhough and Fradley (2005). Then the child began working. Once the child completed the first puzzle (first trial), the RA reset both ball-and-peg sets to the starting and ending state for the next puzzle. Procedures were the same for test trials 2, 3, and 4. The first trial was a 2-move puzzle (could be solved in a minimum of two moves), the second trial was



a 3-move puzzle, the third trial was a 4-move puzzle, and the last trial was a 5-move puzzle. Throughout the test trials, the RA monitored the child's progress for rule breaks; if a child broke a rule, the RA intervened as quickly as possible and had the child correct it. If the child became frustrated or stuck, the RA verbally encouraged the child to continue. If the child continued to display signs of frustration or stated that he/she didn't want to continue, the RA started the trial over again.

There were two outcome measures to describe performance on the TOL task. First, latency to beginning each problem (a measure of planning time) was recorded from the video files. Timing began from the moment the researcher told the child to start, until the first ball cleared a peg in the child's hand. Total planning time in seconds was calculated by adding together scores across all four trials. Second, children got a score that reflected the number of trials they solved using the least number of moves possible. This is known as the total move-value (Fernyhough & Fradley), and is calculated by adding together the number of moves in the correctly solved trials *only* (i.e. those where child solved it in minimum number of *legal* moves), for a total score ranging from 0 to 14 ( $2 + 3 + 4 + 5 = 14$ ). According to Fernyhough and Fradley, "this score is appropriate as it indicates the number of trials correctly solved, while also giving greater credit for solution of the more complex trials" (p. 110). For this task, a legal move occurred when a ball was lifted from a peg and then went back onto a peg some amount (did not have to go all the way down the peg; could go back onto same peg and still count as a move). An illegal move occurred when a ball was lifted from one peg and went onto the table, a

body part other than the hand, or onto a “full” peg, or if a second ball was lifted before placing the first. Planning time and total move-value were used in data analyses.

RAs were trained to do the TOL coding by going through several examples with the lead researcher. Detailed instructions and descriptions were provided for each code category. After training, the RAs coded nine of the same videos and compared results. Discrepancies were discussed and final coding decisions were made. Inter-rater reliability for the two coders was very high for TOL planning time ( $r = .99, p < .05$ ), legal moves ( $r = .998, p < .05$ ), and illegal moves ( $r = .89, p < .05$ ). The rest of the videos were coded by one RA each.

### **Executive Functioning: Inhibitory Control – Replacement of a Dominant Response with a Less-Dominant Response**

The Head-to-Toes task was recently developed by Ponitz and colleagues (2008) and McClelland and colleagues (2007) as an easy and inexpensive way to tap into behavioral inhibition. This task involves the researcher telling the child to touch his/her toes, but the child should actually do the opposite (touch his/her head). It was originally used with children ranging in age from 36 to 66 months (3 – 5.5 years). Children in this age range demonstrated wide variability in task performance (Ponitz et al., 2008). The task has good construct validity, as demonstrated by improvement with developmental age. However, children did not improve with practice, indicating limited to no practice effects. It also demonstrated acceptable convergent validity through small to moderate significant correlations with teacher ratings of behavioral regulation (Ponitz et al., 2008). The task has been translated into Spanish (and back-translated to English), and used with

a native Spanish-speaking sample (Ponitz et al., 2008). However, at around five years of age, several children score at ceiling. Therefore, for the current study, a more complex version was used.

A more complex version known as the Head-Toes-Knees-Shoulders (HTKS) task was developed that builds on the Head-to-Toes task by adding an additional set of instructions where children touch their shoulders when the experimenter says ‘knees’ and vice versa (Matthew, Ponitz, & Morrison, 2009; Ponitz et al., 2009). This task is appropriate for children in early elementary school, as it requires children to hold four rules in mind, as opposed to just the two required in the original Head-to-Toes task (Matthew et al., 2009). Prior research demonstrated that performance on the HTKS task is correlated with parent and teacher ratings of children’s behavior (Ponitz et al., 2009), and with academic achievement (Matthew et al., 2009; Ponitz et al., 2009) in kindergarten. The task has recently been used with first-graders (six-year-olds), but has not yet been used with 7-year-olds (Connor et al., 2010; Wanless et al., 2011). It has also demonstrated cultural validity through recent use with Taiwanese, South Korean, and Chinese samples (Wanless et al., 2011).

For the current study, procedures followed those outlined by Connor et al. (2010). Children were told that they were going to play a game where they would be following some commands from the researcher. First, children were told to touch their head, and then touch their toes. Then the researcher told the child that they were going to be silly and switch the rules. The researcher explained that the child should do the opposite of what the researcher said. So if the researcher said “touch your head,” the child should

actually touch his/her toes, and vice versa. Then the child experienced six practice trials. If the child made a mistake during practice trials, the researcher provided corrective feedback. The first set of 10 test trials followed this. During test trials, the researcher alternated between the two commands (“touch your head” and “touch your toes”) in a predetermined order. Children were expected to follow the “silly rule,” and do the opposite of what the experimenter said. Feedback was not given during test trials.

Next, the researcher moved on to a new set of rules, where the child was told to touch his/her knees and shoulders. The child was first told to touch knees/shoulders in a normal manner. After doing this, the researcher explained the silly rule again. This time, if the researcher said “touch your knees,” the child should touch his/her shoulders, and vice versa. The child then experienced five practice trials with feedback, alternating between the knees/shoulders commands. Finally, the second set of ten test trials began. Before starting, the researcher reminded the child of all four “silly” rules (“When I say touch your head, you touch your toes,” etc.). The researcher alternated between the commands in a predetermined order. Feedback was not provided during test trials.

Following the guidelines of Ponitz et al. (2009), the test trials were scored in the following manner: two points for a completely correct response; one point for a self-corrected response (the child began movement toward the incorrect body part, but then ended by touching the correct body part); zero points for a completely incorrect response. This means that a child could receive a total score ranging from 0 to 40 points. This overall score was used for data analyses. The HTKS has been translated into Spanish (and then back-translated into English), and used with a Spanish-speaking sample (Ponitz

et al., 2009). Variability in scores has been demonstrated by prior research, as well as adequate construct validity when compared with parent and teacher ratings of child behavior (Ponitz et al., 2009).

### **Executive Functioning: Parent Report**

The Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000) was completed by parents as an assessment of children's general EF. The BRIEF is designed for use with children ages 5-18. It is an 86-item questionnaire and rates children's behavior in two domains thought to be critical to EF – behavioral regulation and metacognitive skills (problem solving; Gioia et al., 2000). The BRIEF was originally developed to be a more ecologically valid measure of EF deficits that is easy to complete, requiring only 10 to 15 minutes (Gioia, Espy, & Isquith, 2002; Isquith, Crawford, Espy, & Goia, 2005). Direct assessments administered in a lab setting can elicit a range of EF abilities, but that does not necessarily mean that these patterns manifest in the real world. Questionnaires, on the other hand, tap into real-world manifestations of children's EF by someone who interacts daily with the child. Isquith et al. (2005) asserted that questionnaires should be *complementary* to, and not a substitute for, direct assessment.

The BRIEF requires parents to make responses on a three-point scale (never, sometimes, often) about how often children engage in various behaviors. The questionnaire has 86 items which make up eight subscales: Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor. The first three make up the Behavior Regulation Index (BRI), and the other five

make up the Metacognition Index (MI). Together, the BRI and MI represent the Global Executive Composite (GEC). Higher scores on any scale are indicative of greater impairment in that domain. The BRIEF has demonstrated high internal consistency reliability (.82-.98; Gioia et al., 2000). The BRI, MI, and GEC scores were used in data analyses. The Spanish version of the BRIEF was available if parents chose to complete the questionnaires in Spanish.

### **Externalizing Behavior Problems**

Children's externalizing behavior problems were assessed through a parent survey, the Child Behavior Checklist (CBCL; Achenbach, 1991). This is one of the most widely used assessments of child problem behavior. On his website – the Achenbach System of Empirically Based Assessment (ASEBA) – Achenbach provides a list of over 7,000 publications from around the world that used some version of the CBCL and/or other ASEBA materials (Achenbach, 2010). The CBCL is a 112-item questionnaire that asks parents to indicate how true different statements are about the child (0 = not true; 1 = somewhat or sometimes true; 2 = very true or often true). The statements focus on different problem behaviors the child might engage in. The form is appropriate for use with children 6-18 years of age. There is another version appropriate for use with children younger than 6 years old; however, I only used the form for 6- to 18-year-olds in my study to ensure consistency in terms of the behaviors being rated by different parents. The questionnaire contains nine subscales, which combine together to create an Internalizing Problems score, and Externalizing Problems score, and an Overall Problems score. Both raw and standard scores are available. Larger numbers indicate more

behavior problems. For the current study only results from the Externalizing Problems scale were reported. The CBCL subscales demonstrate good test-retest reliability, ranging from .80 to .94, and good internal consistency with alpha values ranging from .78 to .97. The CBCL has good discriminant validity with the ability to distinguish clinically referred from non-referred children and good content validity through correlations with other measures of child behavior (Achenbach & Rescorla, 2001).

A Spanish version of the CBCL was available so that parents could complete the CBCL in their language of preference (English or Spanish). Prior research has compared the use of the CBCL in different languages and cultures. For instance, Crijnen and colleagues (Crijnen, Achenbach, & Verhulst, 1997, 1999) conducted analyses to compare CBCL results from 12 different countries, including Germany, Greece, China, Belgium, Australia, Sweden, Thailand, Israel, the Netherlands, Jamaica, the U.S., and Puerto Rico. Results indicated that Puerto Rican children had means slightly above the “omnicultural mean,” though this may have been due to sampling (the Puerto Rican sample had a higher completion rate than any other cultural group). Across cultures, externalizing behaviors decreased with age and boys were rated higher than girls on aggression, attention deficits, and delinquency (Crijnen et al., 1997). This research demonstrates the cross-cultural validity of the CBCL.

## RESULTS

### **Exploratory Data Analyses**

Exploratory data analyses were conducted to examine the distribution and potential outliers for all relevant variables, including language variables, behavior problem variables, and EF outcome variables. The distribution for children's English receptive language skills (PPVT percentile) was negatively skewed with many scores in the 80th – 90th percentile. The distribution for children's Spanish receptive skills was fairly normal. For English expressive skills, there were two larger outliers for total words produced in the story, as well as for unique words produced (more than three standard deviations above the mean). For Spanish expressive skills, there was also one large outlier (more than three standard deviations above the mean) for both total words and unique words. However, the primary purpose of the expressive language variables was to help classify bilingual children as DLL vs. balanced and to help create the degree of bilingualism score. Therefore, I decided it was not necessary to remove these outliers, as these variables are not involved in main analyses and outliers did not affect their use in the creation of the other variables. The variable representing degree of bilingualism had a normal distribution with no outliers.

For the EF variables, exploratory data analyses were conducted on go/no-go (GNG) commission errors, HTKS total score, Simon congruent trials reaction time (RT),



Simon incongruent trials RT, Simon Effect score (difference in RT for incongruent vs. congruent trials), DCCS block 3 star errors, TOL planning time, TOL total move-value, and the two BRIEF subscales (behavior regulation and metacognition indices), as well as the BRIEF global executive composite. The distribution for GNG commission errors was very positively skewed, with several scores of 0, 1, or 2 (i.e. many children made none or only a few commission errors), and a few large outliers of 10, 11, 12 (i.e. three children often pressed the spacebar in response to the “no-go” stimuli). Due to their size (two more than three standard deviations and one more than two standard deviations above the mean), analyses were conducted both with and without these outliers. For the HTKS total score, the distribution was negatively skewed with several children having close to perfect scores (scores close to 40), and two outliers on the tail end of the distribution (scores of 18 and 22; two children scored very low overall on the HTKS). Analyses were conducted with and without these outliers.

For the Simon task, the distributions for RT on congruent trials and incongruent trials were quite normal except for one large outlier in each distribution (more than three standard deviations above the mean). Similarly, there were two fairly large outliers (more than two standard deviations above the mean) for the Simon effect score. Analyses were conducted with and without these outliers for all three Simon measures. The distribution was slightly bimodal for DCCS block 3 star errors, with several scores of zero and three (scores ranged from zero to five). For the Tower of London (TOL), the planning time variable had two large outliers. Analyses were conducted with and without these outliers. The distribution for TOL total move-value was fairly normal with no outliers, and had a

slight positive skew (many children had mid to low scores on this variable). The BRIEF subscales and overall score had normal distributions with no outliers. One subject was missing data on the BRIEF (parent left too many items blank). For behavior problems, only the externalizing problems subscale of the CBCL was considered. The distribution for this variable had a slight positive skew (but no outliers), with several parents rating their children very low on behavior problems. Three subjects were missing data on the CBCL due to incomplete parent surveys.

### **Covariate Identification**

Next, I examined relations between the main variables of interest (degree of bilingualism, EF, behavior problems) and demographic variables, including child gender, child ethnicity, child age, total children in the family, annual family income, mother education, father education, marital status, immigrant status, and PPVT score. Child gender was negatively correlated with GNG commission errors ( $r = -.26, p < .05$ ), where girls made fewer commission errors on the GNG (indicating better impulse control). Child age was related to DCCS block 3 star errors ( $r = -.27, p < .05$ ) and GNG commission errors ( $r = -.34, p < .05$ ). In other words, older children made fewer errors on the DCCS and GNG.

Total children in the family was positively related to DCCS block 3 star errors ( $r = .27, p < .05$ ) and negatively related to Simon effect score ( $r = -.30, p < .05$ ), indicating that a larger household was associated with more errors on the DCCS, but a lower Simon effect score (better interference control). Annual family income was not significantly related to any EF or behavior variables. Mother education level was negatively related to

DCCS block 3 star errors ( $r = -.23, p < .05$ ), indicating that children with higher educated mothers made fewer errors on the DCCS. Father education level was not significantly related to any EF or behavior variables.

Marital status (coded as 1 = married/cohabitating and 2 = single/divorced/separated/widowed) was significantly related to the BRIEF behavior regulation index (BRI), metacognition index (MI), and global executive composite (GEC;  $r = -.24, r = -.25$ , and  $r = -.27, p < .05$ , respectively). In other words, children from single-parent homes had fewer EF problems as reported by parents. Immigrant status (native-born = 0, 1<sup>st</sup> or 2<sup>nd</sup> generation immigrant = 1) was negatively related to degree of bilingualism ( $r = -.30, p < .05$ ). Within the group of bilinguals, immigrant children (child or at least one parent born in a country outside the U.S.) were more likely to have a lower score on this variable, indicating more balance between their two languages. This is not surprising, as all but two of the balanced bilinguals were 1<sup>st</sup> or 2<sup>nd</sup> generation immigrants. Finally, PPVT score did not relate significantly to any EF or behavior variables. Due to the large number of potential covariates, only a few were selected to consider in main analyses, including child age, family income, father education, and total children in the family. Child ethnicity was also examined as a second IV in most analyses. These demographics were ones that had the most relations with other main variables of interest and/or that were significantly different across the language groups (see Table 1 and discussion below for language group differences in background variables).

I also examined the inter-correlations between all EF measures. These can be seen in Table 3.

Table 3. Bivariate correlations between all EF outcome measures.

	2	3	4	5	6	7	8	9
1. GNG	-.23 <sup>+</sup>	-.08	-.08	-.09	.19 <sup>+</sup>	.11	.07	.31*
2. HTKS		-.05	-.05	-.04	-.19	.29*	.22 <sup>+</sup>	-.18
3. Cong RT			.89*	-.05	.20 <sup>+</sup>	.08	-.27*	-.02
4. Incong RT				.40*	.18	.14	-.18	-.12
5. Simon Effect					.06	.07	.16	-.24*
6. DCCS Star						-.17	-.07	.15
7. TOL Plan							.08	-.04
8. TOL TMV								-.16
9. BRIEF GEC								

The pattern of correlations changed sufficiently when the outliers for the EF variables (discussed above) were removed, so the correlations reported in Table 3 reflect the removal of the outliers. Specifically, not surprisingly, RT on Simon congruent trials was highly correlated with RT on incongruent trials, meaning that children who had slower RT on one trial type were very likely to also have slower RT on the other trial type (and high scores indicate poorer interference control). The HTKS was positively related to TOL planning time, indicating that children who received higher scores on the HTKS (better inhibitory control) also tended to spend more time planning on the TOL. GNG commission errors were positively related to parent report of EF (BRIEF GEC), indicating that children who performed worse on the GNG (poor inhibitory control) were also rated by parents as having more EF difficulties/deficits. Unexpectedly, the Simon Effect was negatively related to parent report of EF, implying that children with a higher Simon Effect (i.e. less interference control) were rated by their parents as having fewer

EF deficits. Interestingly, RT on Simon congruent trials was negatively related to TOL total-move value (TMV), meaning that children who responded slower on congruent trials were more likely to have a lower total move-value (solved fewer/easier puzzles in least moves possible, as opposed to more/harder puzzles).

There were a few marginal relations of note, as well. GNG commission errors were negatively related to HTKS score and positively related to DCCS star errors, indicating that more errors on the GNG (poorer inhibitory control) were associated with worse performance on both the HTKS (also poorer inhibitory control) and DCCS (poorer cognitive flexibility). RT on congruent trials of the Simon task was positively related to DCCS star errors, indicating that poorer interference control (Simon) was associated with poorer cognitive flexibility (DCCS; larger numbers indicate more errors). Finally, HTKS score was positively related to TOL TMV, indicating that children with higher HTKS scores (better inhibitory control) tended to solve more/harder TOL puzzles in the least moves possible (better planning/problem-solving skills).

### **Language Group Differences in Demographic Variables**

Table 1 presents demographic information for the overall sample, as well as separately by language group. Statistical comparisons were made with language group as both a two-level (monolingual vs. bilingual) and three-level (monolingual vs. DLL vs. balanced bilingual) variable. As seen in Table 1, monolingual and bilingual participants did not differ in child gender, child age, birth order of target child, total number of children in the family, mother education level, marital status, or receptive and expressive English language skills. Monolingual families reported higher annual income ( $M = 7.00$ ,

$SD = 1.00$ ; about \$120,001 - \$175,000) than bilingual families ( $M = 6.04$ ,  $SD = 1.76$ ; about \$80,001 - \$120,000),  $t(72.10) = 3.04$ ,  $p < .05$ . The groups also differed in terms of racial/ethnic composition,  $\chi^2(2) = 29.91$ ,  $p < .05$ . The majority of bilinguals were Hispanic and White. One bilingual was classified as Other. On the other hand, the majority of monolinguals were White, with only four classified as Hispanic or Other. The groups differed in terms of father education, where fathers of monolingual children ( $M = 6.00$ ,  $SD = .75$ ) had higher education than fathers of bilingual children ( $M = 5.39$ ,  $SD = 1.40$ ),  $t(68.66) = 2.47$ ,  $p < .05$ .

Some additional group differences were found when language group was examined as a three-level variable. The only significant findings reported here are those that added to the picture already described for bilinguals and monolinguals. The three groups differed on family income, where monolinguals had significantly higher income than balanced bilinguals,  $F(2, 75) = 4.21$ ,  $p < .05$ . DLLs fell between these two groups, not significantly different from either (see Table 1 for group means and SDs). The groups also differed quite a bit on gender. For monolinguals, the ratio of males to females was close to 50/50 (see Table 1), while only 23.5% of balanced bilinguals were male, compared to 65.5% of DLLs,  $\chi^2(2) = 7.58$ ,  $p < .05$ .

## **Data Reduction**

I attempted to reduce the set of EF outcome variables to a smaller set of factors, first using principal axis factoring, and then using principal components analysis. The various factor analytic models that I attempted had problems for several reasons. Due to small sample size, very different measurement scales, and poor distributions for some of

the measures, the models had difficulty converging on a solution. For models that did converge, the communalities were fairly low (less than .5 for several variables, indicating that the extracted factors were not accounting for much of the variance in the EF variables) and there were several cross-loadings across factors (even after oblique rotation was employed). The loadings on the various factors also did not make theoretical sense. For all of these reasons, EF outcomes were analyzed separately in the main analyses below, as opposed to being combined into fewer factors.

### **Research Question 1**

This question examined group differences in the nine EF variables for the three different language groups (monolinguals, DLLs, and balanced bilinguals). For each EF outcome variable, an analysis of variance (ANOVA) was initially conducted with language group as the IV and the given measure of EF as the DV. Then the selected covariates (age, income, father education, total children) were added into the model separately (ANCOVAs), so as to maintain as large an *N* as possible for each analysis (for instance, a few participants were missing father education data, and a few different participants were missing total children data, so to include both of these as covariates in a model would reduce sample size by even more than just including one covariate). For a given EF variable, any covariates that were significant were then combined together into a final model for that EF variable. The results of the inhibitory control measures (GNG, HTKS, Simon) are presented first, followed by cognitive flexibility (DCCS), then planning/problem-solving (TOL), and finally parent report of EF (BRIEF). When

findings were not significant, the statistics reported are from the simplest model with no covariates. Table 4 presents means and SDs by language group for all EF measures.

Table 4. Means and SDs for EF outcome measures by language group.

	Monolingual ( <i>n</i> = 33)	DLL ( <i>n</i> = 29)	Balanced ( <i>n</i> = 17)
GNG Comm Errors	2.55 (2.44)	2.66 (2.68)	2.18 (2.04)
HTKS Total	33.67 (4.68)	35.31 (3.72)	33.53 (3.86)
Simon Cong RT* <sup>12</sup>	<b>801.30</b> (161.84)	<b>885.50</b> (162.51)	<b>927.90</b> (165.00)
Simon Incong RT	921.43 (255.83)	987.07 (228.02)	971.41 (197.82)
Simon Effect <sup>2</sup>	92.88 (87.55)	76.24 (80.45)	58.65 (81.09)
DCCS Star Errors <sup>3</sup>	1.91 (1.44)	1.55 (1.32)	1.35 (1.37)
TOL Planning* <sup>2</sup>	16.75 (8.87)	<b>19.90</b> (8.66)	12.44 (4.15)
TOL Tot Move-Value	4.82 (3.62)	4.72 (3.03)	6.29 (3.41)
BRIEF GEC	47.79 (8.03)	46.54 (49.59)	49.59 (47.73)

\**p* < .05 (significant group comparisons are bolded)

<sup>1</sup>means are covariate-adjusted means, controlling for age

<sup>2</sup>with outliers removed

<sup>3</sup>group comparison was significant for monolinguals vs. all bilinguals, controlling for age, income, total kids (covariate-adjusted means: monolingual *M* = 2.07, *SD* = 1.30; bilingual *M* = 1.41, *SD* = 1.29)

### Group Differences in Inhibitory Control

To begin, there were no language group (monolingual vs. DLL vs. balanced bilingual) differences in commission errors on the GNG, even after including covariates,  $F(2, 76) = .21, p = .810$ . Results were the same with two large outliers removed. There



were also no group differences (monolingual vs. DLL vs. balanced bilingual) for the HTKS,  $F(2, 76) = 1.51, p = .228$ . This held true even when two large outliers were removed. In other words, as expected, the language groups did not differ on inhibition of a prepotent response (GNG) or on a more motor based inhibitory control task where the individual is required to replace a dominant response with a less-dominant one (HTKS).

Next, I examined interference control via the Simon task. For RT on congruent trials of this task, it first appeared that the groups were not different. However, after removing one large outlier and controlling for age, the language groups did differ significantly, with monolinguals having faster RT than balanced bilinguals,  $F(2, 74) = 3.91, p < .05$ . The DLL group fell in between these two and did not differ significantly from either. See Table 4 for group means. For incongruent trials, the groups did not differ, even after including covariates and removing one larger outlier,  $F(2, 76) = .65, p = .525$ . Both of these findings contradict expectations, as balanced bilinguals were expected to have faster RT on both trial types, an indication of superior interference control.

For the Simon Effect (difference in RT between incongruent and congruent trials), with no covariates in the model, language group did not appear to be significant,  $F(2, 76) = .97, p = .385$ . However, as the group mean for balanced bilinguals ( $M = 58.65, SD = 81.09$ ) was substantially different from that for monolinguals ( $M = 92.89, SD = 87.55$ ) and DLLs ( $M = 95.64, SD = 106.16$ ), an effect size was calculated for this comparison (balanced bilinguals vs. all else), resulting in  $d = .38$ . With such an effect size, it is clear that results are in the expected direction, with balanced bilinguals demonstrating a lower Simon Effect (and hence better interference control) than both DLLs and monolinguals.

A larger sample size for balanced bilinguals would likely lead to significant results. Further, after removing two large outliers from the Simon Effect (both from the DLL group), results were still not significant, but the group means shifted so that DLLs appeared to also have a lower score on the Simon Effect compared to monolinguals (see means and SDs in Table 4). Therefore, I calculated two more effect sizes for the monolingual vs. DLL comparison ( $d = .20$ ) and the DLL vs. balanced comparison ( $d = .22$ ). Interestingly, it appears there was a trend, where DLLs displayed better interference control than monolinguals, and balanced bilinguals were better than DLLs. A larger sample size would be needed to explore this finding further.

### **Group Differences in Cognitive Flexibility**

Next, I examined group differences in star errors on block 3 of the DCCS (i.e. the six trials from this block that had stars on the cards), my measure of cognitive flexibility. Larger scores indicate more errors and hence less cognitive flexibility. Initially, language group was not significant,  $F(2, 76) = 1.04, p = .358$ , even with covariates in the model. However, as there did seem to be a trend, where both DLLs and balanced bilinguals had lower means compared to monolinguals (see Table 4), I chose to collapse the two bilingual groups together and compare them to monolinguals. After controlling for child age, family income, and total children in the family, language group was significant,  $F(1, 68) = 4.30, p < .05$ , where monolingual children ( $M = 2.07, SD = 1.30$ ) made significantly more errors on the DCCS than bilingual children ( $M = 1.41, SD = 1.29$ ). This indicates greater cognitive flexibility for the bilingual group as a whole compared to monolinguals.

### **Group Differences in Planning/Problem-Solving and Parent Report of EF**

For the Tower of London (TOL), I examined group differences for both planning time and total move-value (a score that indicates how many trials of the TOL were solved in the least number of moves possible; the child gets more credit for trials that have a greater number of minimum moves). After removing two large outliers from the planning variable, language group was significant,  $F(2, 74) = 4.46, p < .05$ , where DLLs had significantly more planning time compared to balanced bilinguals. Planning time for monolinguals fell in between these two groups, not significantly different from either (see Table 4 for means). It should be noted, however, that the balanced group had much less variability compared to the other two groups. This finding was unexpected, so I chose to examine it in a little more detail. I discovered that “delay time” (number of seconds from when the RA placed the last ball on the peg sets at the beginning of a TOL puzzle, until he/she told the child to start) was actually significantly lower for the DLL group ( $M = 1.90, SD = 2.21$ ) compared to the balanced bilingual and monolingual groups ( $M = 3.56, SD = 4.29$ ),  $t(73.53) = 2.24, p < .05$ . In other words, the balanced bilingual and monolingual groups unintentionally had significantly more time than the DLL group to examine the TOL puzzles while the RA was still providing instructions. “Planning time” was the number of seconds from when the RA said “start” (at the end of instructions) until the first ball cleared a peg in the child’s hand. It is very likely that the DLL group compensated for this reduced delay time with increased planning time. Therefore, the finding of differences between the groups in planning time is not a real finding, and is instead due to a methodological error.

For TOL total move-value (TMV), there were no significant group differences, with or without covariates in the model,  $F(2, 76) = 1.36, p = .262$ . However, as seen in Table 4, there was an interesting trend, where the balanced group appeared to have a higher mean than the DLL and monolingual groups, with similar variability across the groups. Therefore, I calculated an effect size for the comparison of balanced bilinguals ( $M = 6.29, SD = 3.41$ ) vs. monolinguals/DLLS ( $M = 4.77, SD = 3.33$ ), resulting in  $d = .45$ . Such an effect size implies an interesting trend, where balanced bilingual children appeared to have superior performance on the TOL (solved more/harder puzzles in the fewest moves possible) compared to both DLLs and monolinguals. A larger sample size is needed to further explore this finding. For the last part of RQ1, I examined language group differences in parent report of EF, as measured by the BRIEF. There were no significant group differences for parent report of EF, with and without covariates in the model,  $F(2, 75) = .74, p = .48$ .

## **Research Question 2**

This question examined whether degree of bilingualism related to any measures of EF, which would indicate that, for bilinguals, more or less balance across two languages is related to EF. A series of Pearson correlations was conducted to address this question. No correlations were significant for degree of bilingualism and any EF measure. This held true with and without controlling for the set of background variables described above (age, family income, father education, total children in the family, and ethnicity). However, due to the small sample size ( $N = 46$  bilinguals), I felt it was noteworthy to point out any correlations of .20 or above. After controlling for age, family

income, father education, and total children, there was a small negative relationship between RT on congruent trials of the Simon task and degree of bilingualism ( $r = -.21, p = .214$ ). [It is important to note that some children were excluded from the analysis when controls were included, due to missing data. However, the correlation was similar ( $r = -.22$ ) when controlling just for age, which allowed all bilingual children to be included in the analysis]. In other words, for bilinguals, less balance across their two languages was associated with faster reaction time on congruent trials of the Simon task. This indicates an unexpected trend, where those children who were more proficient in one language over the other were reacting faster on congruent trials.

Due to the large outliers described above for the GNG, HTKS, Simon, and TOL variables, the correlations were conducted a second time with no outliers. This revealed a marginally significant correlation between degree of bilingualism and planning time ( $r = .28, p < .10$ ), where bilinguals with less balance across their two languages also tended to spend more time planning on the TOL. This unexpected finding is in line with the group comparison above in Research Question 1. The correlation was similar after controlling for child age and ethnicity, family income, total children, and father education.

### **Research Question 3**

The purpose of this question was to look for language group (monolinguals, DLLs, and balanced bilinguals) differences in behavior problems. This question was addressed using ANOVAs and ANCOVAs. For this analysis, language group was used as the IV and the standard score from the externalizing subscale of the CBCL served as the DV. The same set of covariates as above was considered (child age and ethnicity, family

income, total children, and father education). With no covariates included in the model, there were no significant group differences in externalizing behavior problems for the three language groups,  $F(2, 73) = 1.03, p = .36$ . Interestingly, after controlling for family income (no other covariates were significant), an unexpected trend emerged, where balanced bilinguals ( $M = 51.53, SE = 2.20$ ) had marginally more parent-reported behavior problems compared to monolinguals ( $M = 45.34, SE = 1.54$ ),  $F(2, 71) = 2.66, p < .10$ . [Means reported here are covariate-adjusted means.] DLLs fell between these two groups ( $M = 48.59, SE = 1.69$ ), not significantly different from either.

#### **Research Question 4**

This question examined the relationship between degree of bilingualism and behavior problems, excluding monolinguals. This question was addressed using Pearson correlations. The same set of demographic variables was considered as controls. Findings revealed no significant correlations between degree of bilingualism and behavior problems, with and without control variables included.

#### **Research Question 5**

This question explored the relations between behavior problems and the EF outcome measures. A series of Pearson correlations were conducted to address this question. The same set of background variables was considered as controls. See Table 5 for all correlations, both with and without outliers removed. Findings revealed a negative correlation between behavior problems and TOL planning time ( $r = -.27, p < .05$ ), and a positive correlation with overall score from the BRIEF ( $r = .51, p < .05$ ). In other words,

children with more externalizing behavior problems (according to parent report) spent less time planning on the TOL and also had more EF difficulties (according to parent report). These correlations held even after controlling for child age and ethnicity, family income, father education, and total children.

Table 5. Correlations between externalizing behavior problems (CBCL) and EF measures, with and without outliers included.

	With Outliers	No Outliers
1. GNG Commission Errors	.001 ( <i>n</i> = 76)	.14 ( <i>n</i> = 74)
2. HTKS Total Score	-.20 <sup>+1</sup> ( <i>n</i> = 75)	-.16 <sup>1</sup> ( <i>n</i> = 73)
3. Simon Congruent Trials RT	-.16 ( <i>n</i> = 76)	-.07 ( <i>n</i> = 75)
4. Simon Incongruent Trials RT	-.20 <sup>+2</sup> ( <i>n</i> = 76)	-.11 <sup>2</sup> ( <i>n</i> = 75)
5. Simon Effect	-.14 ( <i>n</i> = 76)	-.20 <sup>+</sup> ( <i>n</i> = 75)
6. DCCS Star Errors <sup>3</sup>	-.02 ( <i>n</i> = 76)	
7. TOL Planning Time	-.27* ( <i>n</i> = 76)	-.14 ( <i>n</i> = 74)
8. TOL Total Move-Value <sup>3</sup>	-.13 ( <i>n</i> = 76)	
9. BRIEF GEC <sup>3</sup>	.51* ( <i>n</i> = 76)	

\**p* < .05, <sup>+</sup>*p* < .10;

<sup>1</sup>Controlling for family income

<sup>2</sup>Controlling for child age

<sup>3</sup>No outliers to remove

When controlling for child age, behavior problems were marginally related to RT on incongruent trials of the Simon task ( $r = -.20, p < .10$ ), indicating that children with more behavior problems responded faster on these trials. This correlation reduced

somewhat ( $r = -.17$ ) when all controls were included, though that may be partly because more children were excluded from the analysis due to missing data on background variables. When controlling for family income, there was a marginal negative relationship between behavior problems and HTKS score ( $r = -.20$ ,  $p < .10$ ), where children with more behavior problems scored lower (worse performance) on the HTKS task. This correlation reduced somewhat ( $r = -.19$ ) when all controls were considered, though again, this may be due to the reduction in sample size.

Due to large outliers for the GNG, HTKS, Simon, and TOL scores, these correlations were conducted again with outliers removed (see Table 5). The relationship between TOL planning time and child behavior problems decreased and was no longer significant (with no covariates included,  $r = -.14$ ,  $p = .222$ ; with all covariates,  $r = -.18$ ,  $p = .158$ ). Additionally, behavior problems were no longer correlated with RT on Simon incongruent trials (with no controls,  $r = -.10$ ,  $p = .373$ ; controlling for child age,  $r = -.11$ ,  $p = .339$ ) or HTKS score (with no controls,  $r = -.16$ ,  $p = .181$ ; controlling for family income,  $r = -.16$ ,  $p = .192$ ). Unexpectedly, with outliers removed, a marginal relationship was revealed between behavior problems and the Simon Effect ( $r = -.20$ ,  $p < .10$ ), where children with more behavior problems had a lower Simon Effect (which indicates better interference control). This correlation was similar when control variables were included.

## **Research Question 6**

### **EF as Mediator of Language Group – Behavior Problems Relationship**

This last question included a series of mediated and moderated regression models. First, the mediated regression models were analyzed. This model proposed that the EF



constructs would mediate the relationship between language group and behavior problems. In other words, the relationship between language group and behavior problems only exists because of EF (a children's language group – specifically, being bilingual – causes changes in EF, which in turn, leads to changes in behavior). For this model, language group (3 levels; coding described below) was the IV, behavior problems score from the CBCL was the DV, and each relevant EF measure was a potential mediator. The steps necessary to test for mediation as outlined by Baron and Kenny (1986) include first regressing the mediator onto the IV (mediator must be related to IV), then regressing the DV onto the IV (DV must be related to IV), and finally regressing the DV onto both the mediator and IV. From prior analyses, it is already clear that behavior problems (DV) are related to language group (IV), but not to degree of bilingualism. Therefore, only mediation models with language group as the IV (but not degree of bilingualism) were considered. Further, based on the univariate analyses above (language group must be related to EF), the EF measures that were considered as mediators included: Simon congruent trials RT, Simon effect score, DCCS block 3 star errors, TOL planning time, and TOL total move-value (TMV).

To begin, Simon congruency trials RT was considered as the mediator. For this model, language group (3 levels) was represented by two dummy codes with monolinguals as the reference group (one dummy-coded variable compared DLLs (1) to monolinguals (0) and the other compared balanced bilinguals (1) to monolinguals (0)). For Step 1 of mediation, language group predicted a significant portion of the variance in congruent trials RT,  $R^2 = .08$ ,  $F(2, 75) = 3.27$ ,  $p < .05$ . Specifically, the dummy-coded

variable representing DLLs vs. monolinguals significantly predicted RT,  $b = 87.47$ ,  $t = 2.08$ ,  $p < .05$  (DLLs had slower RT than monolinguals), as did the variable representing balanced bilinguals vs. monolinguals,  $b = 108.80$ ,  $t = 2.21$ ,  $p < .05$  (balanced bilinguals also had slower RT than monolinguals). For Step 2 of mediation, after controlling for income, language group predicted a marginally significant portion of the variance in behavior problems,  $\Delta R^2 = .07$ ,  $\Delta F(2, 71) = 2.66$ ,  $p < .10$ . Specifically, after controlling for income, the variable representing balanced bilinguals vs. monolinguals significantly predicted behavior problems,  $b = 6.19$ ,  $t = 2.25$ ,  $p < .05$  (parents reported more behavior problems for balanced bilinguals compared to monolinguals), while the variables comparing DLLs to monolinguals did not,  $b = 3.25$ ,  $t = 1.41$ ,  $p = .164$  (Step 2 is just described once here, as it is the same for all of the following mediation models). For step 3 of mediation, when both language group and Simon congruent trials RT were included together to predict behavior, RT was not a significant predictor, indicating no mediation.

Next, TOL planning time was considered as the mediator. Similar to above, language group (3 levels; two dummy-coded variables with monolinguals as reference group) predicted a significant portion of the variance in TOL planning, where the variable representing balanced bilinguals vs. monolinguals was a marginally significant predictor and the variable representing DLLs vs. monolinguals was not a predictor. However, when including language group and TOL planning time together to predict behavior, TOL planning was not a significant predictor, indicating no mediation.

Simon effect score was entered as the next mediator. However, in the first step of mediation, language group (IV) did not significantly predict Simon effect score. This

mediation model was stopped at this point, as the mediator must be related to the IV in order to test for potential mediation. The next mediator considered was star errors on block 3 of the DCCS. Based on results of univariate analyses, language group was entered as a 2-level variable for this model, comparing all bilinguals to monolinguals. Though language group predicted a significant portion of the variance in DCCS performance and in behavior problems, once the IV and mediator were considered together as predictors, the mediator (DCCS errors) did not significantly predict behavior, indicating no mediation.

Total move-value (TMV) from the TOL was considered as the final mediator. For this model, language group was again entered as a 2-level variable, but comparing balanced bilinguals to DLLs/monolinguals (based on univariate analyses). Again, though language group accounted for a marginal portion of the variance in both TMV and behavior problems, when the mediator and IV were considered together as predictors (and controlling for income), the mediator did not significantly predict behavior problems. Overall, there was no evidence to support the theoretical model of EF as a mediator of the relationship between language group and behavior problems.

### **Language Group as Moderator of EF – Behavior Problems Relationship**

The next set of regressions examined a moderation model exploring whether language group (levels/coding explained below) moderated the relationship between EF and behavior problems. In other words, does the relationship between various EF constructs and behavior problems differ for different language groups? Based on Vygotskian socio-cultural research that indicates that the use of language as a self-

regulatory mechanism reorganizes the brain, and based on the plethora of research on bilingualism and EF, I theorized that knowledge of two languages might reorganize EF for bilinguals. In turn, EF might be related differently to other areas of functioning for bilinguals vs. monolinguals. Working from this idea, I then conducted bivariate correlations to help guide my selection of EF measures and language group comparisons for the moderation analyses. I began by conducting bivariate correlations between all EF measures and the CBCL externalizing score, separately for monolinguals and bilinguals, and then for DLLs only and balanced bilinguals only (see Table 6 for all bivariate correlations by language group). I visually compared the size and direction of the correlations across groups to see what EF measures might be related to behavior problems differently based on language group. From this, I chose to further explore the following EF measures in moderated regression analyses: GNG commission errors, HTKS total score, Simon congruent RT and incongruent RT, DCCS star errors, TOL total move-value (TMV), BRIEF behavior regulation index (BRI), and BRIEF metacognition index (MI). This visual comparison of bivariate correlations also helped me determine how to code language group for each moderated regression model – I chose to compare the groups that appeared to have different patterns of relationships. For instance, as seen in Table 6, the relationship between Simon congruent (and incongruent) trials RT and parent-reported behavior problems appears to be different for monolinguals vs. bilinguals as a whole group. Therefore, when examining this EF measure, I would choose to code “language group” as monolingual vs. bilingual.

Table 6. Bivariate correlations between externalizing behavior problems score and all EF outcome measures, separately by language group.

	Monolingual	Bilingual	DLLs	Balanced
Direct EF Measures				
GNG Com Errors	.12 ( <i>n</i> = 32)	.15 ( <i>n</i> = 42)	.01 ( <i>n</i> = 25)	.36 ( <i>n</i> = 17)
HTKS Total Score	-.16 ( <i>n</i> = 32)	-.19 ( <i>n</i> = 42)	-.10 ( <i>n</i> = 26)	-.38 ( <i>n</i> = 16)
Simon Congruent RT	-.33 <sup>+</sup> ( <i>n</i> = 32)	.01 ( <i>n</i> = 43)	.04 ( <i>n</i> = 26)	-.05 ( <i>n</i> = 17)
Simon Incongruent RT	-.33 <sup>+</sup> ( <i>n</i> = 32)	-.04 ( <i>n</i> = 43)	.04 ( <i>n</i> = 26)	-.15 ( <i>n</i> = 17)
Simon Effect	-.19 ( <i>n</i> = 33)	-.19 ( <i>n</i> = 42)	-.13 ( <i>n</i> = 25)	-.24 ( <i>n</i> = 17)
DCCS Blk 3 Star Errors	-.06 ( <i>n</i> = 33)	.08 ( <i>n</i> = 43)	-.07 ( <i>n</i> = 26)	.30 ( <i>n</i> = 17)
TOL Planning Time	-.11 ( <i>n</i> = 32)	-.18 ( <i>n</i> = 42)	-.11 ( <i>n</i> = 26)	-.21 ( <i>n</i> = 16)
TOL Total Move-Value	-.13 ( <i>n</i> = 33)	-.16 ( <i>n</i> = 43)	-.09 ( <i>n</i> = 26)	-.31 ( <i>n</i> = 17)
Parent Report of EF				
BRIEF Behav Reg Index	.47* ( <i>n</i> = 33)	.62* ( <i>n</i> = 43)	.61* ( <i>n</i> = 26)	.63* ( <i>n</i> = 17)
BRIEF Metacog Index	.50* ( <i>n</i> = 33)	.43* ( <i>n</i> = 43)	.09 ( <i>n</i> = 26)	.75* ( <i>n</i> = 17)
BRIEF Global Exec Comp	.57* ( <i>n</i> = 33)	.48* ( <i>n</i> = 43)	.18 ( <i>n</i> = 26)	.77* ( <i>n</i> = 17)

\**p* < .05, <sup>+</sup> *p* < .10

All moderation analyses were conducted first with no control variables, and then with child age and family income entered in the first step as controls (separately and together). I chose to consistently include these controls as they were frequently related to other variables in univariate analyses, and because they preserved the largest sample size (no children missing age data and only one missing income data).

To begin, language group was examined as a moderator of the relationship between inhibition of a prepotent response (GNG commission errors) and behavior problems (CBCL externalizing score). I chose to code language group as balanced bilinguals (1) vs. DLLs/monolinguals (0) because the relationship between GNG and CBCL seen in Table 6 appeared to be similar for DLLs and monolinguals, but different for balanced bilinguals. GNG commission errors and language group were entered in Step 1 of the regression, and the interaction term was entered in Step 2. There was no evidence for moderation. Despite this, it is interesting to note that the relationship between commission errors (a measure of inhibition) and behavior problems appears to be much stronger for balanced bilinguals compared to DLLs and monolinguals (see Table 6). It may be that another variable not considered here is driving this pattern, and/or that a larger sample size is needed to further explore this trend.

Next, I explored language group as a possible moderator of the relationship between inhibitory control (HTKS total score) and behavior problems. I again chose to code language group as balanced bilinguals (1) vs. DLLs/monolinguals (0) because the relationship between HTKS and CBCL scores appeared to be different for balanced bilinguals compared to the other two language groups (see Table 6). The best final model (see Table 7) included income as a control in Step 1, followed by language group and HTKS score in Step 2, and the interaction term in Step 3. Results indicated that the interaction between language group and HTKS predicted a marginally significant portion of the variance, above and beyond its component parts and family income,  $\Delta R^2 = .04$ ,  $\Delta F(1, 68) = 2.83$ ,  $p = .097$ ,  $b = -1.55$ ,  $t = -1.68$ ,  $p = .097$ .

Table 7. Moderated regression results for language group as a moderator of the relationship between EF (two types of inhibitory control) and behavior problems.

	IV = HTKS Score				IV = Simon Cong RT			
	<i>B</i>	<i>SE(B)</i>	$\beta$	$\Delta R^2$	<i>B</i>	<i>SE(B)</i>	$\beta$	$\Delta R^2$
STEP 1:								
Family income	1.60*	.74	.248	.062*	1.21 <sup>+</sup>	.68	.205	.042 <sup>+</sup>
STEP 2:				.055				.048
Family income	1.79*	.74	.278		1.54*	.72	.260	
IV	-.36	.28	-.145		-.004	.01	-.071	
Lang Group <sup>1</sup>	3.95	2.54	.179		4.15 <sup>+</sup>	2.17	.234	
STEP 3:				.035 <sup>+</sup>				.041 <sup>+</sup>
Family income	1.96*	.74	.304		1.70*	.72	.287	
IV	-.20	.29	-.081		-.019 <sup>+</sup>	.01	-.374	
Lang Group	57.28 <sup>+</sup>	31.81	2.592		-15.35	10.98	-.867	
Interaction: IV*LangGrp	-1.55 <sup>+</sup>	.92	-2.413		.02 <sup>+</sup>	.01	1.244	

\* $p < .05$ , <sup>+</sup> $p < .10$

<sup>1</sup>Lang group for HTKS = balanced vs. DLL/monolingual; for Simon = bilingual vs. monolingual

As seen in Figure 2, the relationship between inhibitory control (HTKS score) and externalizing behavior problems (CBCL score) was stronger for balanced bilinguals compared to DLLs/monolinguals. For balanced bilinguals, as HTKS score increased (indicating better behavioral inhibition), CBCL score decreased (indicating fewer behavior problems) at a faster rate than for the other language groups. Results held (and barely changed; b-value increased to -1.59) after also controlling for child age. Results with only income as a control are reported for parsimony.

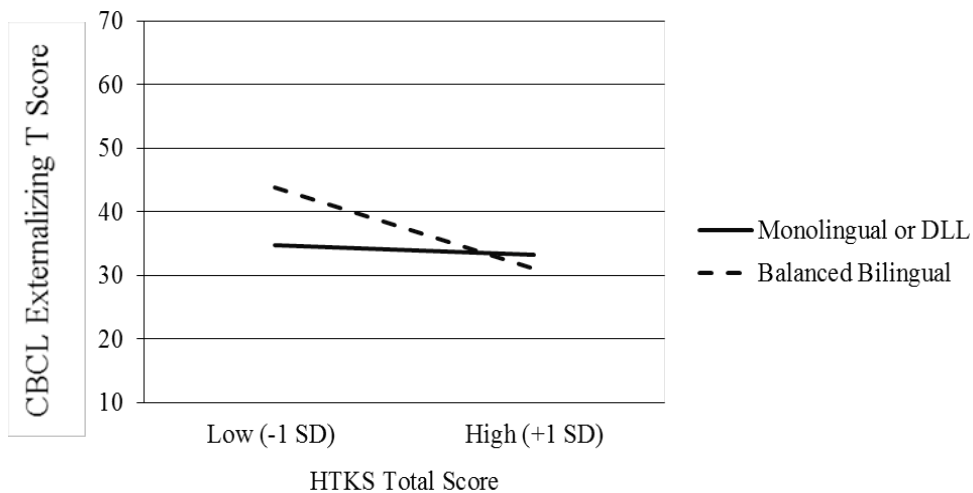


Figure 2. Language group as a moderator of the relationship between inhibitory control (HTKS score) and externalizing behavior problems (CBCL).

I next examined whether language group moderated the relationship between Simon congruent trials RT (an indicator of interference control) and behavior problems (CBCL externalizing score). This time, language group was coded as all bilinguals (1) vs. monolinguals (0), as the bivariate relationship between Simon congruent RT and CBCL appeared to be different for monolinguals compared to bilinguals as a whole (see table 6). The best final model (see Table 7) included family income in Step 1, language group and Simon congruent RT in Step 2, and the interaction between group and RT in Step 3. Results revealed a marginally significant interaction,  $b = .02$ ,  $t = 1.81$ ,  $p = .074$ , that predicted a unique portion of the variance in behavior problems,  $\Delta R^2 = .04$ ,  $\Delta F(1, 69) = 3.28$ ,  $p < .10$ . For monolingual children, the relationship between Simon congruent trials RT and CBCL behavior problems was stronger than for bilingual children. In other words, within the group of the monolingual children, those who were slower to respond on congruent trials (poorer interference control) had fewer behavior problems as reported



by parents, while this relationship did not appear to exist for bilingual children (see Figure 3). Results held (and did not change) when also controlling for child age.

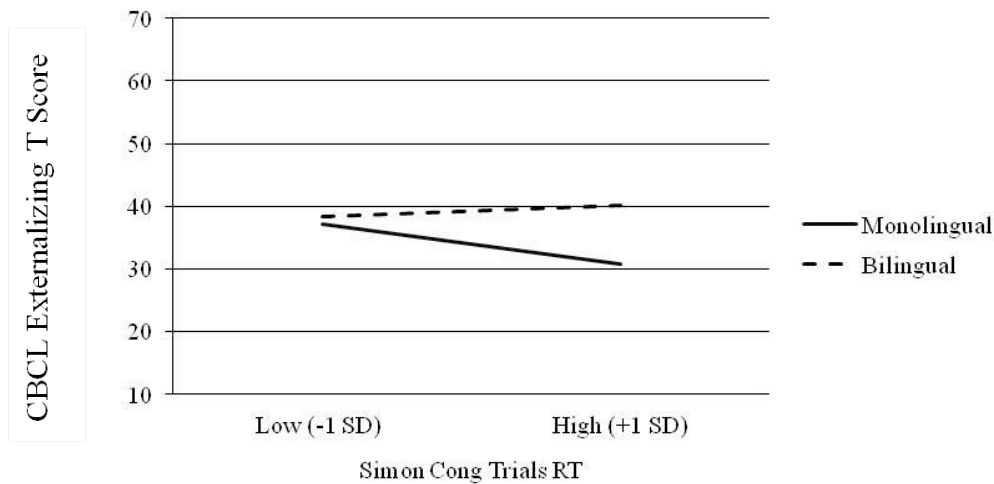


Figure 3. Language group as a moderator of the relationship between interference control and externalizing behavior problems.

Next I examined whether the relationship between Simon incongruent trials RT and behavior problems was moderated by language group (bilingual vs. monolingual). There was no evidence of moderation. Despite this, it is interesting to note that the correlation between Simon incongruent RT and behavior problems was much stronger (-.33) for monolinguals vs. bilinguals (close to 0), which was the same pattern found for congruent trials. As noted above, it may be that another variable not considered is driving this pattern, and/or that a larger sample size is needed to further explore this trend.

Next, language group was examined as a moderator of the relationship between DCCS star errors (cognitive flexibility) and behavior problems. Language group was

coded as balanced bilinguals vs. DLLs/monolinguals, as the relationship between DCCS and behavior problems appeared to be much stronger for balanced bilinguals compared to other language groups. I entered language group and DCCS star errors in Step 1 of the regression, and the interaction term in Step 2. There was no evidence for moderation. However, it is interesting to note that the bivariate relationship between cognitive flexibility and behavior problems was much stronger for balanced bilinguals than for the other two language groups (see Table 6). As already mentioned, some other, unexamined variable might account for this trend, and/or a bigger sample size may be needed. I also examined whether language group moderated the relationship between TOL total move-value (TMV; problem-solving) and behavior problems. I entered TOL TMV and language group (balanced vs. DLLs/monolinguals, as the pattern seemed to differ across these two groups) in Step 1 of the regression, followed by the interaction in Step 2. Results indicated no evidence of moderation. Again, it should be noted that the bivariate pattern between TOL TMV and behavior pattern was quite different for balanced bilinguals compared to the other language groups (see Table 6).

Finally, language group was examined as a moderator of the relationship between parent-reported EF (BRIEF) and behavior problems (CBCL; also parent report). It is important to note that relationships observed in these final moderated regressions may be over estimates due to single-source bias. As seen in Table 6, bivariate patterns were different for the behavior regulation index (BRI) vs. the metacognition index (MI), so these two subscales were examined in separate moderated regressions. It was not necessary to also examine the overall Global Executive Composite (GEC) score because

it is directly computed from the BRI and MI scales. As a reminder to the reader, the BRIEF is intended to tap into “everyday” aspects of EF; how EF manifests in the child’s everyday environment (Gioia et al., 2002; Isquith et al., 2005). More specifically, the BRI is intended to measure children’s inhibitory skills, ability to shift attention, and emotional control. The MI is intended to assess children’s working memory, initiation skills, ability to plan/organize, and ability to monitor one’s own behavior.

First, language group (bilingual vs. monolingual, due to the different bivariate correlations observed for bilinguals as a whole compared to monolinguals) was examined as a moderator of the BRI and behavior problems relationship. There was no evidence of moderation, with and without controls in the model. I also tested whether language group (DLLs vs. balanced bilinguals/monolinguals) moderated the relationship between the BRIEF MI and behavior problems (see Table 8). I chose to examine language group in this manner as the bivariate pattern for DLLs was very different from that for balanced bilinguals and monolinguals (see Table 6).

Table 8. Moderated regression results for language group as a moderator of the relationship between parent report of EF (BRIEF MI) and behavior problems.

	IV = BRIEF MI			
	<i>B</i>	<i>SE(B)</i>	$\beta$	$\Delta R^2$
STEP 1:				.210*
IV	.460*	.11	.458	
Lang Group <sup>1</sup>	1.39	1.94	.075	
STEP 2:				.045*
IV	.59*	.12	.584	
Lang Group <sup>1</sup>	24.36*	11.21	1.31	
Interaction: IV*LangGrp	-.49*	.23	-1.25	

\*p < .05

<sup>1</sup>Lang group = DLLs vs. balanced bilinguals/monolinguals

In Step 2, the interaction term predicted unique variance in behavior problems, above and beyond its component parts,  $\Delta R^2 = .05$ ,  $\Delta F(1, 72) = 4.32$ ,  $p < .05$ ,  $b = -.49$ ,  $t = -2.08$ ,  $p < .05$  (see Table 8). In other words, the relationship between parent-reported metacognitive skills and parent-reported behavior problems was weaker for DLLs than for balanced bilinguals and monolinguals. For balanced bilinguals and monolinguals, children with higher scores on MI (indicating more deficits in metacognition) also had more parent-reported behavior problems (see Figure 4). However, for DLLs, the relationship was close to zero; parent report of MI did not appear to relate to parent report of behavior problems for this group. This finding held (with a slightly higher b value of  $-.51$ ) after controlling for age and income in Step 1. As the b-values are very similar, the results of the simplest model (no controls) are reported in Table 8 for parsimony.

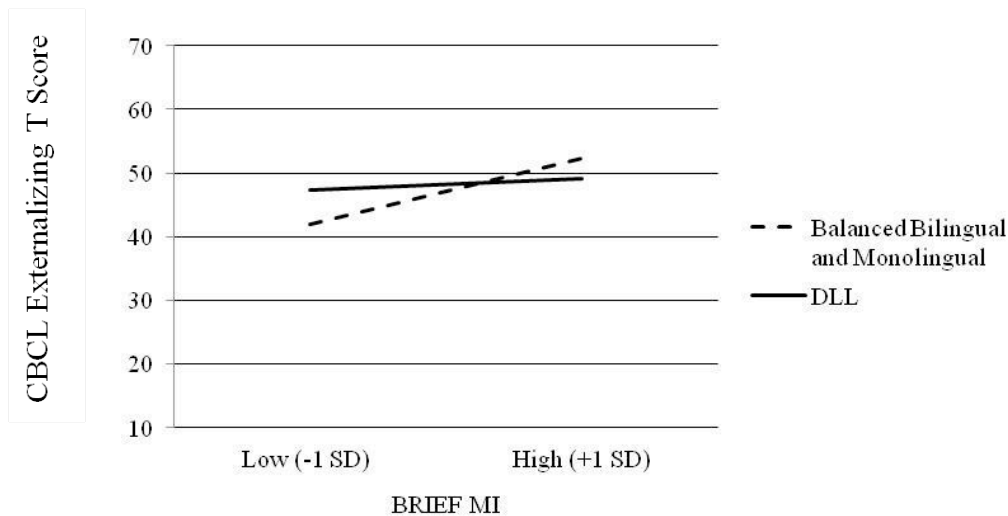


Figure 4. Language group as a moderator of the relationship between metacognitive functioning (BRIEF) and externalizing behavior problems (CBCL).

### **Degree of Bilingualism as Moderator of EF – Behavior Problems Relationship**

I was also interested in whether the relationship between EF and behavior problems might differ by degree of bilingualism as a continuous variable. This area is one with little to no prior research, so these moderated analyses were more exploratory in nature, to examine the utility of conceptualizing “bilingualism” as a continuous variable, as opposed to just a yes/no distinction. These analyses excluded monolinguals, as they do not have a score for degree of bilingualism. Following the same steps above, I first examined whether the bivariate relationships differed for two different levels of the moderator. As this moderator was continuous, I created a median split, with half of the bilinguals (those above the median) receiving a 1 and the other half (those below the median) receiving a 0. The 1’s were considered to be low on degree of bilingualism (less balanced across their two languages; had higher scores on the degree of bilingualism variable) and the 0’s were considered to be high on degree of bilingualism (more balanced across the two languages; had lower scores on the degree of bilingualism variable). Then I conducted bivariate correlations between all relevant EF measures and CBCL externalizing score, separately for the “low” and “high” bilingual groups. See Table 9 for these correlations. From this, I chose to further explore the following EF measures: GNG commission errors, TOL planning time, and BRIEF metacognition index (MI).

Table 9. Bivariate correlations between externalizing behavior problems score and all EF measures, separately for those “high” and “low” on degree of bilingualism.

	“High” on Degree of Bilingualism	“Low” on Degree of Bilingualism
Direct EF Measures		
GNG Com Errors	.32 ( <i>n</i> = 21)	-.03 ( <i>n</i> = 21)
HTKS Total Score	-.18 ( <i>n</i> = 21)	-.29 ( <i>n</i> = 21)
Simon Congruent RT	.06 ( <i>n</i> = 22)	-.02 ( <i>n</i> = 21)
Simon Incongruent RT	-.04 ( <i>n</i> = 22)	-.03 ( <i>n</i> = 21)
Simon Effect	-.21 ( <i>n</i> = 22)	-.19 ( <i>n</i> = 20)
DCCS Blk 3 Star Errors	.15 ( <i>n</i> = 22)	-.03 ( <i>n</i> = 21)
TOL Planning Time	-.38 <sup>+</sup> ( <i>n</i> = 21)	-.17 ( <i>n</i> = 21)
TOL Total Move-Value	-.12 ( <i>n</i> = 22)	-.20 ( <i>n</i> = 21)
Parent Report of EF		
BRIEF Behav Reg Index	.68* ( <i>n</i> = 22)	.56* ( <i>n</i> = 21)
BRIEF Metacog Index	.66* ( <i>n</i> = 22)	.07 ( <i>n</i> = 21)
BRIEF Global Exec Comp	.73* ( <i>n</i> = 22)	.09 ( <i>n</i> = 21)

\**p* < .05, <sup>+</sup> *p* < .10

First, I examined whether degree of bilingualism moderated the relationship between GNG commission errors (inhibition of a prepotent response) and CBCL externalizing score (behavior problems). The best, final model (presented in Table 10) included gender as a control in Step 1, degree of bilingualism and GNG commission errors in Step 2, and the interaction term in Step 3. (Gender was included as a control

because it had a significant bivariate relationship with commission errors, but with no other EF measures).

Table 10. Moderated regression results for degree of bilingualism as a moderator of the relationship between inhibition of a prepotent response (GNG) and behavior problems (CBCL).

IV = GNG Commission Errors				
	<i>B</i>	<i>SE(B)</i>	$\beta$	$\Delta R^2$
STEP 1:				.002
Gender	.79	2.74	.045	
STEP 2:				.032
Gender	1.96	3.00	.113	
IV	.67	.65	.170	
Degree of Bilingualism	.88	1.94	.076	
STEP 3:				.102*
Gender	4.45	3.12	.257	
IV	.89	.63	.225	
Degree of Bilingualism	6.09 <sup>+</sup>	3.11	.526	
IV*Degree Bilingualism	-1.81*	.86	-.536	

\* $p < .05$ , <sup>+</sup>  $p < .10$

Results revealed that, after controlling for gender, the interaction term significantly predicted behavior problems,  $b = -1.81$ ,  $t = -2.09$ ,  $p < .05$ , accounting for a unique portion of the variance,  $\Delta R^2 = .10$ ,  $\Delta F(1, 37) = 4.38$ ,  $p < .05$ . GNG commission errors were positively related to behavior problems (as errors increased, behavior problems also increased) for bilinguals who were more balanced across their two languages, but commission errors were negatively related or not related to behavior problems at all for bilinguals who were less balanced across their languages (see Figure 5). In other words, it seemed that for bilingual children who are more fluent in both of

their languages, direct assessment of inhibitory control was related positively to parent report of behavior (where more problems in one were related to more problems in the other). However, for bilingual children who tend to be stronger in one of their languages over the other, direct assessment of inhibitory control was not related to parent report of behavior. This finding held (and stayed about the same,  $b$  value = -1.87) when age and income were also included as controls in Step 1; results with only gender as a control are reported for parsimony (the interaction was not significant without gender as a control).

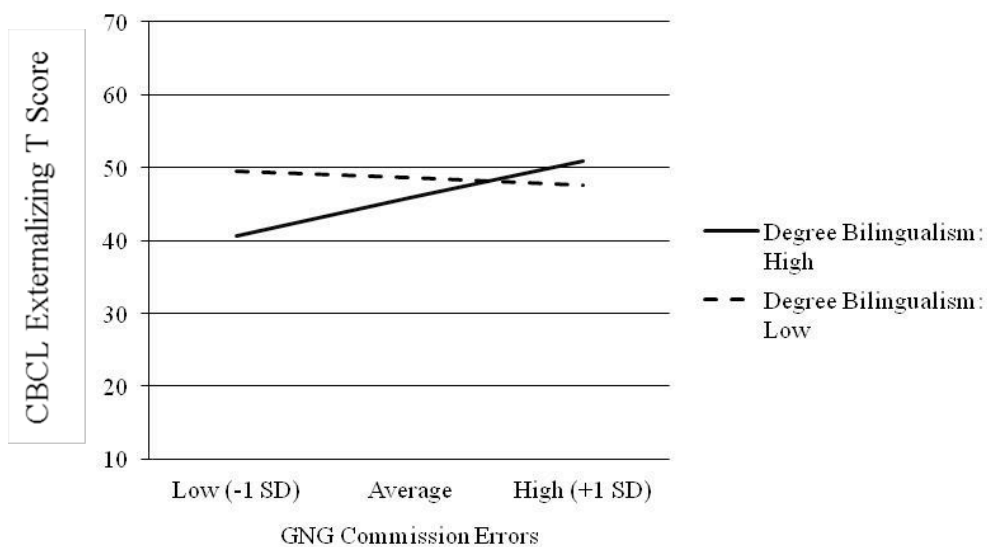


Figure 5. Degree of bilingualism as a moderator of the relationship between inhibition of a prepotent response (GNG commission errors) and externalizing behavior problems (CBCL).

Next, I considered whether degree of bilingualism moderated the relationship between TOL planning time and behavior problems. There was no evidence of moderation. Finally, I examined degree of bilingualism as a moderator of the relation



between BRIEF metacognition (MI) scores and behavior problems. This pattern of findings was less clear. First, age was entered as a control in Step 1, BRIEF MI scores and degree of bilingualism were entered in Step 2, and the interaction term was entered in Step 3. This resulted in significant moderation, where the interaction term predicted unique variance in behavior problems,  $\Delta R^2 = .09$ ,  $\Delta F(1, 38) = 4.56$ ,  $p < .05$ ,  $b = -.40$ ,  $t = -2.14$ ,  $p < .05$  (this predictor was only marginally significant when child age was not included as a control). In other words, MI scores from the BRIEF were strongly related to behavior problems (as MI problems increased, so did behavior problems) for bilinguals who were more balanced in their two languages, whereas MI skills were not really related to behavior problems for less balanced bilinguals (see Figure 6). However, when family income was also included as a control in Step 1, the interaction term was no longer significant,  $b = -.31$ ,  $t = -1.67$ ,  $p = .10$ .

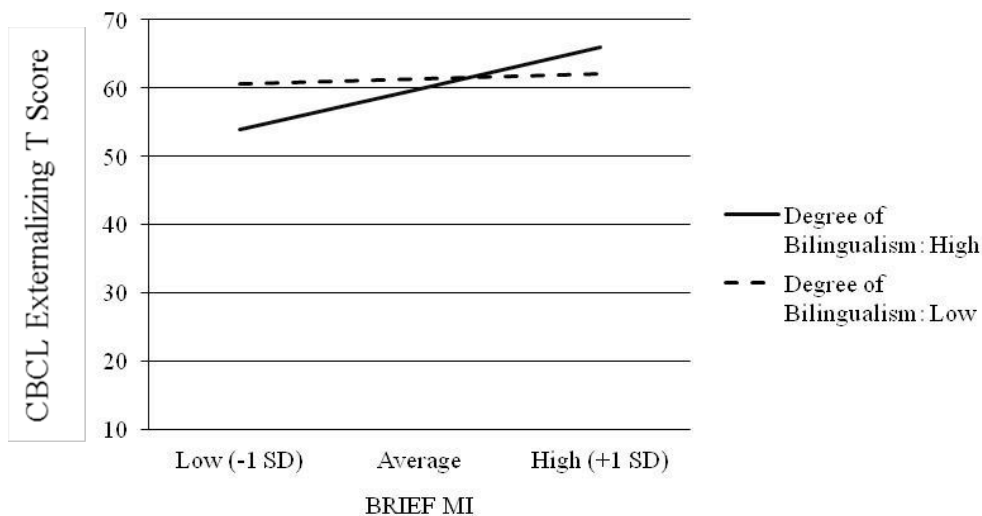


Figure 6. Degree of bilingualism as a moderator of the relationship between metacognitive skills (BRIEF) and externalizing behavior problems (CBCL).

## DISCUSSION

The purpose of the current study was to examine relations between bilingualism, executive functioning (EF), and behavior problems for a sample of bilingual and monolingual 5- to 7-year-old children. The main study goals included: 1) to add to the current body of literature on relations between bilingualism and EF, 2) to further this literature by examining a bilingual group with little prior research (Spanish-English bilinguals), as well as an EF component not explored much with bilinguals (planning/problem-solving skills), 3) to further explore potential EF advantages for a group of children not fully fluent in their second language (Spanish-English Dual Language Learners; DLLs), 4) to add to the current body of literature on the relations between EF and externalizing behavior problems, and 5) to explore potential mediating and moderating relations between bilingualism, EF, and behavior problems.

### **General Findings on EF Development and Theory**

Before examining main study findings in further detail, general findings related to EF will be discussed, including inter-relations among EF measures, implications for the latent structure of EF, age-related trends in EF development, and potential ceiling and floor effects with the current sample. First of all, several of the EF measures were related to one another in expected ways. Children with less ability to inhibit a prepotent response (i.e. suppress a dominant response when cued to do so) had more difficulty with other

forms of impulse control (i.e. inhibition of a dominant response and replacing it with a less-dominant one), as well as poorer cognitive flexibility. Children with better impulse control tended to use more planning time on problem-solving tasks, and also demonstrated more accurate problem-solving overall. Longer reaction time (RT) on congruent trials of the Simon task, potentially an indicator of poorer interference control (Lu & Proctor, 2001), was related to poorer cognitive flexibility, as well as less accurate problem-solving. There was some evidence for the validity of the parent rating of EF, as it correlated positively with commission errors on the GNG: more deficits in EF as measured by parent report were related to more deficits in inhibitory control as measured by direct assessment. However, parent report of fewer EF deficits was unexpectedly related to poorer interference control. The reason for this last finding is unclear.

Second, the various EF components assessed in the current study were somewhat correlated with one another, but not that highly, implying that not just one construct was being assessed. These findings support the theory of EF as one construct with multiple inter-related components proposed by researchers such as Anderson (2002) and Miyake and colleagues (2000). Unexpectedly, unlike prior research, the EF variables in the current study were not easily reduced into a fewer number of latent factors. This is likely due in part to poor distributions on some of the variables, as well as small sample size. In Carlson and Meltzoff's (2008) study, they submitted their EF variables to a principal components analysis with varimax rotation and a sample size of 50, and the model converged, resulting in two factors. It appears that their EF variables were clearly tapping into two separate components (tasks involving cognitive conflict that needed to be

resolved, similar to the Simon task in the current study, and tasks involving behavioral inhibition or delay). My tasks, on the other hand, were intended to assess a wide variety of EF components (three different types of inhibitory control; cognitive flexibility; and planning/problem-solving). If at least four separate (but inter-related) EF components were truly being assessed, then each latent factor would have had only a few indicators, meaning that this theoretical model was not really appropriate for factor analysis (it would need more indicators per factor and/or fewer hypothesized factors).

Next, I examined developmental trends in EF by correlating child age with each EF measure. In terms of inhibitory control, child age was significantly related to commission (pushing button in response to no-go trial) and omission errors (pushing button in response to go trial) on the go/no-go task, where older children made fewer errors, indicating improvements with age in the ability to inhibit a prepotent response, as well as in attention. This is consistent with prior research (Brocki & Bohlin, 2004). Surprisingly, age was *not* related to performance on the HTKS, meaning that older children did not necessarily perform better than younger children on this inhibitory task that involved more motor movement. This is most likely because several children received very high scores on this task, no matter their age. This finding is not consistent with prior research, which has found age-related improvement on this task from preschool up through age 6.5 (Ponitz et al., 2008). Age was also related to improvement in accuracy for both trial types (congruent and incongruent) on the Simon task, but was not related to reaction time (RT). This was surprising, as others have found age-related improvement in RT, indicating an increase in interference control with age (Rueda et al.,

2005). As expected from prior research (Best & Miller, 2010; Carlson 2005), older children made fewer errors on block 3 of the DCCS, indicating age-related improvement in cognitive flexibility. Finally, on the Tower of London, older children made fewer rule breaks and illegal moves than younger children, and spent less time overall completing each trial, indicating improvements with age in planning and problem-solving. This is also consistent with prior research (Luciana & Nelson, 1998).

In terms of age-appropriateness of the selected EF measures, it appears that none of the measures demonstrated floor effects, with the possible exception of the hard TOL problems. In other words, when looking at performance of 5-year-olds in the sample (the youngest age group), children demonstrated good variability on the GNG, HTKS, Simon, DCCS, and TOL tasks. At least a few children made perfect scores on all tasks, except for the TOL. When examining the performance of just older children (7-year-olds), there was good variability in performance on the TOL, and Simon reaction time. There was decent variability on the HTKS – only one child received a perfect score, but most children still scored very highly. However, for the GNG and DCCS, there were potential ceiling effects; these tasks may have been too easy for older children. On the GNG, most children made zero, one, or two commission errors, and all but four made no omission errors. This was likely due to characteristics of the task (i.e. interval between trials, length of time trial was on screen, stimuli used for “go” and “no-go”). For the DCCS block 3 performance, all but three of the older children made one or no star errors (possible range was 0 to 6), indicating that this task may have been too easy for older children. Zelazo (2006) indicated that this advanced version of the DCCS was appropriate for children *up*

to age 7. It was unclear if the task had ever been used with children over the age of 7 (i.e. 7.5 years), but it appears that it is a bit too easy for these children. These relations with age, or lack thereof, and ceiling effect(s) may have contributed to the failure of the factor analyses in supporting multiple interpretable factors of EF.

### **Findings on Second Language Fluency and EF**

Some support was found for the current hypotheses, and in line with prior research, regarding the relationship between bilingualism and EF. As expected, there were no language group differences in simple, more response-based forms of inhibitory control. This included inhibition of a prepotent response (measured by the GNG task), and inhibition involving a choice between competing responses, with no conflict present (measured by the HTKS task). This is consistent with prior research comparing bilinguals and monolinguals on inhibitory control tasks that do not require attentional control, such as the Day-Night Stroop-like task (Martin-Rhee & Bialystok, 2008), which has similar inhibitory demands as the HTKS, and “delay” tasks (Carlson & Meltzoff, 2008). Researchers believe this is because the practice bilingual children get at monitoring and controlling two competing languages does not translate well to situations that do not require control of attention (Martin-Rhee & Bialystok, 2008). With the GNG and HTKS, the child is presented with a single piece of information and must choose between competing *responses*, not something with which they have extra linguistic practice over monolinguals. It is important to remember that part of the reason for the use of the HTKS in the current study was to include a measure of inhibition that required more motor movement on the part of the child. There is some research that indicates that children in

the U.S. who are bilingual or learning English as a second language (many of whom are Hispanic/Latino) have better social skills and fewer behavior problems compared to monolingual English-Speaking peers, at least during early-middle childhood (De Feyter & Winsler, 2009; Hair et al., 2006; Han, 2010; Han & Huang, 2010). If these children really do have advanced control over their behavior compared to peers, it is still possible that bilingual children might score higher than monolinguals on a task like the HTKS, for a reason *other than dual language practice* (i.e. maybe due to other cultural factors). A sample with different language/ethnic groups and a sample that varies more in behavior problems would be needed to test this theory (see below for discussion on limited range of behavior problems for the current sample).

Unexpectedly, the language groups did *not* differ in reaction time (RT) for incongruent trials of the Simon task, and monolinguals were actually *faster* than balanced bilinguals on congruent trials (DLLs fell in between). This is not consistent with prior research where balanced bilinguals were faster than other language groups on *both* trial types of the Simon task (Bialystok et al., 2004; Martin-Rhee & Bialystok, 2008).

According to the theory of bilingual advantage, bilingual children get a lot of practice with control of attention and inhibition (Bialystok, 2001). Both languages are “activated” at all times in the bilingual brain, and the mechanisms used to suppress the unneeded language at any point in time are thought to be the same as those used during inhibition that involves control of attention to multiple pieces of information (Green, 1998).

There are a few potential reasons for the null findings. First, Bialystok (2009) has pointed out that the parameters of the Simon task are critical for demonstrating the

bilingual advantage. The task must present enough cognitive conflict; if it does not, the task may be too easy, and bilingual children will likely not differ from monolingual children in their RT performance. In three prior studies (two with 4-year-olds, one with 8-year-olds), Martin-Rhee and Bialystok (2008) set up the Simon task so that half of the trials were congruent, and half were incongruent. In the current study, 60% of trials (34 trials) were congruent. This decision was made as a compromise between research by Martin-Rhee and Bialystok, and findings by Borgmann and colleagues (2007) that a proportion of congruent trials closer to 75% is necessary for a more reliable Simon Effect (i.e. slowed RT on incongruent vs. congruent trials) to occur. However, it may be that changing the congruent to incongruent trials from a ratio of 1:1 up to a ratio of 3:2 made the task *too* hard, thereby diminishing or even eliminating the bilingual advantage. Further, Martin-Rhee and Bialystok only had 48 trials maximum (even with eight-year-olds), where the current study had 60. So another possibility is that, instead of (or in addition to) being too hard, children may have simply gotten bored and started responding slower overall (though not at the expense of accuracy, as accuracy was fairly high across all language groups).

Despite the unexpected findings with RT on the Simon task, there did appear to be a sizeable group difference in the Simon Effect score (difference in reaction time for incongruent vs. congruent trials). Though not statistically significant, balanced bilinguals had a substantially lower score for the Simon Effect compared to DLLs, and DLLs had a lower score than monolinguals. I computed effects sizes, and they were small to moderate ( $d = .40$  for balanced bilinguals vs. monolinguals;  $d = .22$  for balanced bilinguals vs.



DLLs;  $d = .20$  for DLLs vs. monolinguals), which is particularly impressive given the small sample size. The lack of significant findings for the Simon Effect, despite large group differences in favor of balanced bilinguals, has also been demonstrated in prior research (Martin-Rhee & Bialystok, 2008). This is potentially due to small sample sizes for both the current and prior studies. The resulting effect sizes in the current study indicated a bilingual advantage for both balanced bilinguals *and* DLLs.

These results have a few important implications. First, it seems that balanced bilingual children can still display an advantage in EF, even when residing in a country where there is not much cultural or educational support in place for minority languages. Second, the advantage in interference control for DLLs over monolinguals is an important new finding. To my knowledge, no other studies have found advantages in EF for children considered to be still learning their second language (“unbalanced” bilinguals). In their study, Carlson and Meltzoff (2008) included a sample of English-dominant DLLs that had six months of exposure *on average* to Spanish through a Spanish-English immersion school (but only exposed to English at home). Their DLL group was not different from the monolingual group on measures tapping into interference control (though their balanced bilingual group did perform better). It is possible that exposure for six months to a second language, for only two or three hours each day, five days per week, is not enough time for effects on EF development to be measurable.

In the current study, DLL children were exposed to their second language anywhere from a minimum of six months (only one participant) up to their entire lives.

This finding implies that, with enough exposure to a second language (and setting/context of exposure might also matter), positive benefits can be seen for non-linguistic domains of cognitive development, even when the child is not fully proficient in the second language. Bialystok and Majumder (1998) also included a group of Bengali-English DLLs in their study and found no differences from monolinguals on tasks requiring attentional control (while a third group of balanced bilinguals did perform better). In that study of 7- to 9-year-olds, the DLLs attended school all in English, and were considered to be English-dominant, but had both Bengali and English exposure in the home. Presumably, this exposure was from birth, so there must be something more than just length of exposure that is critical for the even small EF advantages, like those demonstrated by the DLLs in the current study. Bialystok and Majumder concluded that part of the reason for these children's limited proficiency in Bengali was because they received no formal instruction in this language.

Next, when examined as a three-level variable, the language groups did not differ significantly in cognitive flexibility (DCCS star errors). However, when bilinguals as a whole were compared to monolinguals (language group as a two-level variable), and after controlling for child age, family income, and total children in the family, bilinguals *did* have significantly fewer errors on the DCCS compared to monolinguals. This finding implies an advantage in cognitive flexibility for both balanced bilinguals *and* DLLs over monolinguals. The balanced bilingual advantage in cognitive flexibility is consistent with other studies using the DCCS (or similar) measures (Bialystok, 1999; Bialystok & Martin, 2004; Carlson & Meltzoff, 2008; Okanda et al., 2010). However, the advantage

in cognitive flexibility demonstrated for not completely bilingual DLLs in the current study has not been documented in prior research. A similar trend was found in Carlson and Meltzoff's study, where their balanced group outperformed their monolingual group, and their DLL group fell in between these two, not significantly different from either. Some other potential reasons why the DLL group displayed EF advantages in interference control and cognitive flexibility are discussed further below.

Interestingly, balanced bilinguals also appeared to have an advantage in planning/problem-solving skills, an EF component that has been examined very little in prior research with bilinguals. For the Tower of London (TOL) total move-value (TMV; a sum of the total number of moves used to solve each of the four TOL puzzles, only for those solved in the least number of moves), there were no significant group differences. However, as the means appeared to be pretty different in size for balanced bilinguals compared to the other groups, I calculated an effect size. The resulting effect was moderate in size,  $d = .45$ . In other words, balanced bilinguals solved more/harder puzzles in the least number of moves compared to the other groups.

It appears that balanced bilinguals had some advantage in planning/problem-solving, though the sample size was likely too small to reach statistical significance. These findings support those of the only other study (Moreno et al., 2010; an unpublished poster presentation) that could be found that examined the TOL with balanced bilinguals, where a bilingual advantage was also demonstrated (though it is not clear what score was used to measure TOL performance). It is not clear why the DLL group did not demonstrate some small advantage in this EF component, when they did demonstrate an

advantage in interference control and cognitive flexibility. In another study by Luciana and Nelson (2002), no group differences were found in TOL performance for DLLs compared to monolinguals, though the language experience and proficiency of the DLL group in that study were very unclear (so they may have in fact been balanced bilinguals). It may be that, because planning/problem-solving is a more complex EF that builds on other components (i.e. working memory, inhibitory control, cognitive flexibility, etc.), it has a higher threshold of bilingualism for advantages to be conferred. I recommend that planning/problem-solving skills for both balanced bilinguals and DLLs be explored further in future studies with larger sample sizes.

Unexpectedly, DLLs were actually significantly different from balanced bilinguals and monolinguals on the other measure from the TOL, where DLLs had significantly more planning time (total seconds from when the research assistant told the child to “start” until the first ball cleared a peg in the child’s hand) compared to the other two groups. Upon further exploration, it seems that this group difference was actually the result of a methodological flaw, as opposed to being a true group difference.

Unintentionally, DLLs had significantly less “delay time” on the TOL than the other two groups. As a reminder, “delay time” arose from a methodological flaw, where sometimes RAs continued giving directions after completing the set up of a given TOL problem, allowing the child extra time to study the puzzle before the RA actually told the child to begin working (RAs were supposed to say “start” to the child immediately upon placing the last ball on the peg sets). Most likely, because DLLs had significantly less “delay time” than the other groups, they spent more time “planning” as a result.

The current study had several direct assessments of various EF components. However, it is possible that one-time assessments in a lab setting might not accurately reflect children's EF strengths and weaknesses. Some researchers have pointed out the need for parental report of a particular behavior as a complement to direct assessment of that behavior (Isquith et al., 2005). Therefore, the current study also included parent report of children's EF using a measure (the Behavior Rating Inventory of Executive Function; BRIEF) that was intended to tap into EF deficits as they would manifest in children's everyday environment (Gioia et al., 2002; Isquith et al., 2005). Interestingly, the language groups did not differ on this measure. There are a few potential explanations for this null finding. First, the BRIEF was designed to tap into EF *deficits*. Indeed, this is demonstrated by the fairly strong positive correlation found between BRIEF scores and CBCL behavior problem scores, which supports the extensive literature on the link between externalizing behavior problems and EF deficits (Barkley et al., 1992; Pennington & Ozonoff, 1996; Raaijmakers et al., 2008; Willcutt et al., 2005). However, the theory on the bilingual advantage does not posit any EF deficits for particular groups; rather, typically developing bilinguals are advanced in certain areas of EF compared to typically developing monolingual peers.

Second, because the BRIEF taps into so many components of EF, it may be that those components with a demonstrated bilingual advantage are not isolated well enough on this parent report measure to reveal any group differences. Third, there was substantial cultural variability in the current sample. It is possible that parents from different language and/or cultural backgrounds have different expectations and standards for their

children's behavior. The BRIEF rating scale is based on frequency (never, sometimes, often) and includes items such as "mood changes frequently," "does not think before doing," and "is impulsive." What a U.S.-born White parent views as "often" for behaviors having to do with impulse or emotional control, for instance, might be at a different threshold than what parents from different cultural backgrounds view as "often" for these same behaviors. Scores on the BRIEF did not differ by ethnic group or language in which the parent completed the survey (Spanish or English), but there may be some other unmeasured cultural variable that relates to the null findings for the BRIEF.

In addition to being examined categorically, second language fluency was also examined as a continuous variable (with only balanced bilinguals and DLLs receiving scores for this variable). Surprisingly, "degree of bilingualism" (how balanced the two languages were for a child) was not significantly related to any measures of EF. In other words, gradual changes in "how bilingual" a child was did not relate significantly to any changes in EF. As no other studies have attempted to examine bilingualism on a continuum, it is unclear why the findings were not significant. It may be that effects are small, and a larger sample size is needed for detection. It may also be that this variable I created did not reflect what I intended it to. As a reminder, the "degree of bilingualism" variable was an average of four difference scores for children's Spanish and English receptive skills, expressive skills, parent report of proficiency, and language use (the scores were standardized before being averaged, as they were all on different scales). Perhaps this variable should have been weighted, with certain aspects of language

background receiving more weight than others. For instance, Morton and Harper (2007) said that actual language *use* by bilingual children should be considered most important.

### **Findings on Second Language Fluency and Behavior**

Unexpectedly, after controlling for family income, balanced bilinguals had marginally more behavior problems than monolinguals, according to parent report. DLLs fell in between, not significantly different from either group. This went against the hypothesis; I expected balanced bilinguals to have the least behavior problems out of the three language groups. While some research has demonstrated that bilingual children demonstrate better social skills and fewer behavior problems (Hair et al., 2006; Han, 2010; Han & Huang, 2010), other research is less clear on this topic (Preciado et al., 2009). Furthermore, it should be noted that the finding was only marginal, and there was restricted range on behavior problems for the current sample (discussed further below in Limitations), where most children were rated fairly low by parents on this variable. Therefore, to say that balanced bilinguals had marginally more behavior problems than another group does not really mean that these children were high on externalizing problems. This relationship needs to be examined further in studies with samples with noticeably higher behavior problems. As with the BRIEF, CBCL scores did not differ by ethnic group or language of the survey (Spanish vs. English), but some other unmeasured cultural variable may be related to these scores. Further, “degree of bilingualism” was not significantly related to behavior problems. This may be due to the reasons already noted above regarding this variable. It may also be because “how bilingual” a child is does not actually relate to behavior problems. If bilingual children really do have better self-

control skills compared to monolingual peers (which, again, would need to be investigated in future studies), this may be due to other cultural or family background variables, as opposed to knowledge of two languages.

### **Findings on EF and Behavior**

I hypothesized that behavior problems would be related to EF in such a way that increases in externalizing behavior would be associated with decreases (or worse performance) in EF. Behavior problems were only significantly related to a few measures of EF. Unexpectedly, a lower score for the Simon Effect was marginally related to more behavior problems. In other words, children with better interference control tended to be rated by parents as higher on externalizing behavior problems (I expected children with better interference control to be rated as *lower* on behavior problems). It is unclear why this pattern emerged. It is likely due to the fact that balanced bilinguals were rated by parents as slightly higher in behavior problems compared to the other language groups (they were expected to be *lower* in problem behaviors), and balanced bilinguals also had better interference control (as expected), so a spurious correlation emerged between behavior problems and interference control. Behavior problems were also related to parent report of EF, where children with more behavior problems were rated by parents as having more EF difficulties. This finding supports my hypothesis, as well as the vast literature demonstrating a link between externalizing behavior problems and EF deficits (Barkley et al., 1992; Pennington & Ozonoff, 1996; Raaijmakers et al., 2008; Willcutt et al., 2005). However, it must be noted that both of these measures are parent report, so single-source bias is likely a problem. This is a documented issue in social science



research, where correlations between variables from a single source (same parent), even if those variables are intended to measure *different* behaviors, are often higher than correlations of variables from different sources/methods intended to measure the *same* behavior (Campbell & Fiske, 1959).

### **Findings on the Interplay of Language, EF, and Behavior**

I hypothesized that EF would mediate the relationship between language group/bilingualism and behavior problems, where being bilingual would lead to changes in EF, which would, in turn, lead to changes in externalizing behavior. Results did not support my hypothesis; there was no evidence that any measure of EF mediated the relationship between language group and behavior. When externalizing behavior problems are evident for a given child, their presence often has to do with problems with inhibition/impulse control (Pennington & Ozonoff, 1996). Indeed, research on EF for kids with behavior problems shows that inhibition is the EF component most consistently impaired for these children (Pennington & Ozonoff, 1996). And based on prior research, as well as results from the current study, bilingual children do not necessarily display an advantage in EF for the kinds of inhibition that are typically impaired for children with behavior problems (Carlson & Meltzoff, 2008; Martin-Rhee & Bialystok, 2008). If children from immigrant and language-minority families really do demonstrate better social skills and fewer behavior problems compared to their native English-speaking peers, perhaps some other factor mediates the relationship between language group and behavior, such as parenting, family values, or other cultural factors. However, it is important to keep in mind that the current sample did not display very high behavior problems (discussed further in

Limitations section below). This sample is not similar to other studies of non-clinical samples with elevated behavior problems. It is important for future research to investigate this link again with a new sample before more firm conclusions are drawn about the relations between second language fluency, EF, and behavior.

I also hypothesized that language group would moderate the relationship between EF and behavior problems, where there would be a stronger relationship for children who are DLLs or balanced bilinguals (as EF performance increased, behavior problems would decrease at a faster rate; or as EF *problems* increased, behavior problems would *increase* at a faster rate) compared to monolinguals. This hypothesis was partially supported. For inhibitory control as measured by the HTKS, there was a stronger negative relationship (controlling for family income) between this EF and behavior problems for balanced bilinguals (though not for DLLs) compared to the other language groups. This same trend, though non-significant, was observed for the relationship between inhibition of a prepotent response (GNG task) and externalizing behavior, as well as for that between cognitive flexibility (DCCS) and externalizing behavior, and between TOL performance and behavior problems. When bilingualism was considered as a continuous variable (“degree of bilingualism”), the relationship between inhibition of a prepotent response and behavior problems (controlling for gender) was stronger for bilinguals with more balance across their two languages (as problems with inhibitory control increased, behavior problems increased at a faster rate) compared to those with less balance. This is consistent with the non-significant trend described above.

In contradiction to this hypothesis, however, the relationship between RT on Simon congruent trials and behavior problems was actually stronger for *monolinguals* compared to bilinguals; as RT decreased, behavior problems increased at a faster rate for monolinguals. The pattern of moderation was the same for incongruent trials, though not significant. It is unclear why this finding emerged. Perhaps monolingual children with poorer impulse control were simply reacting faster on this task. Also in contradiction to the hypothesis, the relation between parent-report of metacognition (MI) and behavior problems was stronger for balanced bilinguals and monolinguals compared to DLLs; as MI problems increased, behavior problems also increased at a faster rate for balanced bilinguals and monolinguals, but did not appear to be related for DLLs. This same pattern was found when bilingualism was examined as a continuous variable, where the relationship between MI and behavior was stronger for bilinguals with more balance across their two languages compared to those with less balance. It is also unclear why this pattern emerged. However, it is important to note that both of these measures were parent-report, so findings may be over-estimated due to single-source bias (discussed further in Limitations section below).

Taken together, these findings seem to indicate that the relationship between EF and behavior problems may be different for balanced bilinguals vs. other language groups. Perhaps, as theorized in the Introduction, bilingualism does change the structure of EF for these children, and that affects the relations EF has with other behaviors/areas of functioning. There has not been much research regarding the prefrontal architecture of language development and how it might influence EF development. However, there is a

small cognitive neuroscience literature that indicates that multiple brain areas, both within and outside of the prefrontal cortex, are involved in EF components like inhibitory and attentional control (Casey et al., 2000; Durston et al., 2003; Simmonds et al., 2007). Future work in this area could explore other EF tasks and how exposure to two languages might change prefrontal architecture over time. Future research would need to further investigate these findings, as firm conclusions cannot be drawn based on the results from this one study with a small sample size.

### **Strengths and Contributions of the Current Study**

The methodology and findings from the current study contribute to the literature in several important ways. First, most of the children in my DLL group were exposed to their second language for much longer than six months (the average length of time DLLs were exposed to L2 in Carlson & Meltzoff, 2008). Also, some of my DLL children came from native Spanish-speaking families (where at least one parent spoke Spanish as a native language) as opposed to all English-speaking families. Some went to English-only schools and some went to Spanish-English immersion schools. Therefore, I can tentatively conclude that my DLL group better represents the experience and demographics of many Hispanic children in the U.S., who have regular exposure to a non-English language at home (sometimes in addition to English) and exposure to primarily English at school (except for those children who happen to attend one of the few Spanish-English immersion schools in the U.S.). [As noted below in the Limitations section, this variability can also be a drawback].

Second, many researchers have noted the difficulty in defining “bilingualism” as a construct (Baker, 2006). It is multi-faceted, including such factors as length of exposure, timing of exposure (i.e. from birth or later in life), receptive and expressive proficiency in each language, frequency of use of each language, and the actual contexts in which each is used (i.e. home vs. school vs. community) (Baker, 2006). Despite the multi-pronged nature of the construct, several prior studies only assess one language instead of both (Bialystok & Feng, 2009; Bialystok & Shapero, 2005; Bialystok & Viswanathan, 2009), and/or assess only receptive skills and not expressive skills (Bialystok & Feng, 2009; Bialystok & Majumder, 1998; Bialystok & Shapero, 2005; Bialystok & Viswanathan, 2009; Martin-Rhee & Bialystok, 2008). The current study attempted to improve upon this gap by gathering direct assessments of receptive *and* expressive language skills in *both* languages, in addition to parent report of proficiency, frequency, and context and length of exposure for each language. This information was used to help classify children into language groups.

Third, due to the complexity of the construct, researchers have also pointed out that bilingualism should be thought of on a continuum (Baker, 2006). However, for several practical reasons, most studies operationalize “bilingualism” as a yes/no variable (Bialystok, 2001). To my knowledge, the current study is the first to attempt to create a continuum of bilingualism, by calculating a “degree of bilingualism” variable, based on as much quantifiable language background information as possible (receptive skills, expressive skills, parent report of proficiency, and parent report of language use/context). This allowed for analyses to be conducted that could examine the *linear* relation between

bilingualism and other variables (EF, behavior problems), as opposed to just mean differences between groups.

Fourth, much prior research on the link between bilingualism and EF has been conducted in countries where multiple languages are supported and appreciated (e.g., English paired with other languages in Canada; Bialystok & Feng, 2009; Bialystok & Majumder, 1998; Bialystok & Shapero, 2005; Martin-Rhee & Bialystok, 2008). It is noteworthy that, in a country where one language is clearly the majority, and where children from diverse language home environments might not have as much cultural support or community acceptance for a second minority language (and may even be stigmatized for speaking it), bilingual children who are fluent in two languages are still demonstrating advantages in EF over monolinguals. Research in this area should continue with samples of bilingual children considered more “at-risk” (i.e. from low-income families; with mental health or behavioral issues) to further tease apart what child, family, and cultural factors might be related, or not related, to the well-established bilingual advantage in cognitive development.

Other strengths of the current study included examination of an under-studied group of DLLs and balanced bilinguals (Spanish-English), as well as an under-studied age range (5 to 7). Others have pointed out the need for more research on this dual language group (Carlson & Meltzoff, 2008), as well as on the EF development of the age range just beyond preschool (Best et al., 2009). I also added new tasks that have not been examined much in terms of relations between bilingualism and EF (HTKS, TOL).

Finally, I conducted mediating and moderating models to explore potential relations between bilingualism, EF, and behavior problems, a unique feature of this study.

### **Limitations and Considerations for Future Research**

As with any empirical investigation, the current study had several limitations that should be noted. First, the sample size was rather small ( $N = 79$ ), especially when broken down by language group (balanced bilingual  $n = 17$ , DLL  $n = 29$ , monolingual  $n = 33$ ). There were also several pairs of siblings included in analyses. Removal of one sibling per pair revealed a similar pattern of results (group  $N$ 's changed as follows: balanced bilingual = 14, DLL = 26, monolingual = 28), so the siblings were left in to maintain sample size and power. Despite this, the sample size of the current study is similar to (and sometimes larger than) other recent studies also examining the bilingual advantage for EF (Carlson & Meltzoff, 2008; Martin-Rhee & Bialystok, 2008). For instance, in three studies by Martin-Rhee and Bialystok, with children ranging in age from four to eight years, sample sizes were 34 (bilingual  $n = 17$ ), 41 (bilingual  $n = 17$ ), and 32 (bilingual  $n = 13$ ), respectively. In another study by Carlson and Meltzoff, sample size was 50 with only 12 balanced bilinguals and 21 DLLs. What's more, these studies all found significant differences between the bilingual and monolingual groups, in favor of bilinguals, on tasks involving cognitive/attentional control (i.e. interference control). These studies, as well as findings from the current study, suggest that the bilingual advantage is a robust enough effect that it can be detected even with small sample sizes and limited statistical power.

Within the DLL group for the current study, more than half of participants were male (65.5%), but within the balanced bilingual group, the majority (76.5%) were female. Therefore, for the group of bilinguals as a whole, gender was potentially confounded with language group. However, in terms of EF measures, gender was only significantly related to GNG commission errors. Gender was included as a control variable in analyses that involved GNG scores. Further analyses should attempt to tease out the possible influence of gender on relations between language group, EF, and behavior.

Three children (two in the DLL group and in the balanced group) had substantial exposure to a third language other than English or Spanish. It is possible that EF patterns would be different for such children. However, parents of these children rated children's understanding/ speaking of the other language as "poor" to "ok," so it is unlikely that any of the children would be considered fluent in the third language. Furthermore, in Martin-Rhee and Bialystok's study (2008, Study 3), their monolingual English-speaking group of children was exposed to Hebrew for about 2 hours every day at school, yet the authors still considered this group monolingual and still found EF differences when comparing this group to a fully balanced bilingual group. The authors discussed "the importance of being functionally bilingual for cognitive effects to emerge" (Martin-Rhee & Bialystok, 2008, pg 91). In other words, the authors stressed that bilinguals must *use* both languages on a pretty regular basis to see some effect on EF.

The language groups were not matched on income (monolingual families reported higher income than balanced bilingual families), though even the group with the lowest income still had a fairly high average (between \$60,000 and \$120,000). This variable was



used as a consistent covariate in analyses to help account for this disparity. However, there may be other factors related to family SES that are also related to child cognitive development that were not controlled for in the current study. Despite this, according to Bialystok (2009), matched income across groups is not necessarily a requirement for detecting group differences in EF because SES and bilingualism are also thought to have independent effects on EF development. In other words, in theory, bilingual children from lower SES backgrounds could still demonstrate the bilingual advantage, even when compared to monolinguals from more advantaged backgrounds.

Due to the study location, the majority of the sample not only had very high income (average between \$60,000 and \$80,000), but several of the parents had higher education as well (most had at least a Bachelor's degree). Therefore, the current results may not be generalizable to other samples of Spanish-English bilinguals around the country, as it is well known that a large percentage of Hispanic/Latino families in the U.S. are from lower-SES backgrounds (U.S. Census Bureau, n.d.). It is important to replicate this kind of study with lower-income samples of Spanish-English bilinguals. In Carlson and Meltzoff's (2008) study, their Spanish-English bilingual sample had substantially lower family income and parent education compared to their DLL and monolingual groups, yet the bilingual advantage was still demonstrated for tasks involving interference control. This provides some indication that, as Bialystok (2009) noted, bilingualism seems to have an influence on EF that is independent from SES, when other relevant group differences are taken into account (e.g., verbal ability, age, etc.). However, as the current study is the first to also examine relations with behavior

problems, it is important to replicate this study with a sample of balanced bilingual and DLL children that more accurately reflect the economic characteristics of the majority of Hispanic/Latino families in the U.S.

As with many quasi-experimental designs, the current study is not necessarily representative of the larger population, as random sampling was not used. I attempted to recruit families from lower SES, however, it turned out that most families were from higher SES. Further, the families that participated are likely ones where the parents are very interested in children's language development and bilingualism, and/or in promoting research on child development in general. Such parents might be more involved in their children's lives, may place more focus on appreciation for cultural diversity, etc., resulting in potential differences between study children and the general population of 5- to 7-year-olds. It is important for future studies to seek out participants that are less likely to volunteer as readily, such as families in poverty.

One important goal of the current study was to examine children's behavior problems, and how these problems related to EF and language. Unfortunately, due to limitations with recruitment methods, the measure of behavior problems suffered from range restriction: the large majority of children were rated by their parents as quite low in behavior problems overall (for the CBCL externalizing scale, the potential range for the standard score was 33 to 100, with higher scores indicating more problems, but the actual range for the current sample was 33 to 67). Only six children fell in the "at risk" or "clinical" range (according to cut-offs created for the CBCL), putting them at potential risk for a disorder associated with externalizing behavior problems, such as ADHD (and

only one participant was actually diagnosed with ADHD, and one had been previously evaluated for ADHD). Additionally, behavior problems were only measured via parent report, and parents may over- or under-estimate their children's behavior problems. Despite this, there was still some variability in behavior problem scores (sample  $M = 47.74$ ,  $SD = 8.87$ ). It is important to replicate this work with samples of children that demonstrate more variability in behavior, as well as to measure behavior problems using alternative methods or raters (i.e. teacher report or observational methods to supplement parent report).

For many practical reasons, several research assistants helped collect study data. Though all RAs received extensive training on the measures and were required to use scripts during assessments (organized by task in a spiral notebook), there remains the possibility that task administration may have varied across RAs. This is especially an important consideration for the measures that were more dependent on consistency in RA language, including the Tower of London, the Dimensional Change Card Sort, and the Head-Toes-Knees-Shoulders task. For instance, as noted above, the significant group difference in planning time, where DLLs had more planning time compared to the other language groups, is very likely a direct result of the decreased "delay time" for this group compared to the others. As a reminder, "delay time" arose from a methodological flaw on the part of RAs, where sometimes RAs unintentionally provided children with extra time to study a given TOL puzzle before beginning. Interestingly, DLLs had significantly less delay time than the other groups, and spent more time "planning" as a result.

As another example, for one subject, when the RA was administering the first part of the HTKS (trials with only “touch your head” or “touch your toes” commands), the RA forgot to give corrective feedback during practice trials when the child made mistakes. As a result, the child scored very low on the first ten test trials. However, for the second half of the task (when commands for knees and shoulders were added in), corrective feedback *was* given on practice trials, and the child scored much higher on the second set of ten test trials. Overall, the child’s HTKS score was very low, and this was one of the subjects excluded from analyses as an outlier. (However, this mistake on the part of the RA was very uncommon, and no other children’s scores appeared to be affected).

My samples of balanced bilingual and DLL children were highly variable, in that they came from very different cultural backgrounds. Seventy-eight percent of the bilingual group was comprised of either 1st or 2nd generation immigrants, from 24 different countries. Bialystok (2009) points out how the bilingual and monolingual groups in her studies are very similar in terms of cultural and economic background because they lived in the same neighborhoods and went to the same schools, and most were born in the same country (Canada). It is possible that my lack of significant findings for my balanced bilingual group may be partially attributable to this variability, especially for tasks that have historically demonstrated an advantage for balanced children over DLLs and monolinguals (Simon task – interference control; DCCS – cognitive flexibility). Additionally, this variability, might also account for, or be related to, some of the EF advantages demonstrated for my DLL group.

Finally, as mentioned in the Participants section, the majority of children in my DLL group were English-dominant instead of the intended characteristic of Spanish-dominance. This was unexpected, and primarily due to sampling issues (limited recruitment locations/methods for bilingual and DLL children). Despite this, findings for the DLL group are still very relevant, as the characteristic of English-dominance within the child, despite a native Spanish-speaking parent, is a very common situation for many children in Spanish-speaking homes in the U.S., especially after they progress a few years through public schooling. So this aspect of the study can also be viewed as a strength (discussed above), as this is a group of DLLs and balanced bilinguals without much prior research, but that exist in large numbers in the U.S. However, future studies should focus on recruiting a sample of Spanish-dominant DLLs, and/or balanced bilinguals where English is not spoken much at home. It is important to determine if the bilingual advantage is present for low-income children living in a culture where their dominant L1 in the home is a minority language.

## **General Conclusions and Implications**

Findings from the current study have several important, practical implications for both practice and policy. First, as noted, the majority of DLLs in the current study were English-dominant. Such children may face unique obstacles within U.S. society, compared to children who are Spanish-dominant (i.e., true “English Language Learners”). Anecdotal evidence from the current study (during casual conversations with parents) suggests that in homes where one or both parents are native Spanish speakers, parents sometimes focus on English language skills with children at the expense of

Spanish skills. Many parents said it was just “easier” to use English, since the children learned through that language in school. Current study results have important implications for such families in terms of supporting Spanish language development. It is indeed important to promote English, but not at the expense of the Spanish language, as knowledge and use of two languages seems to confer cognitive advantages for children.

Additionally, several Spanish-speaking parents of English-dominant DLLs (and even a few balanced bilinguals) reported that their children seemed to be rejecting the Spanish language, and did not really like speaking it at home, even though they had the ability to do so. Indeed, during the Spanish story-telling task, a few children refused to produce much (if any) Spanish, though their parents assured the RA that the children spoke it at home (and children’s vocabulary scores reflected working knowledge of Spanish for these children). This may be because children in all-English public schools feel peer pressure, even at this young age, to speak English.

Interestingly, several DLLs and balanced bilinguals were attending Spanish-English immersion schools in the local area. Some of these children were from English monolingual homes, and some had at least one Spanish-speaking parent. In contrast to bilinguals at English-only schools, several parents of children attending an immersion school reported that their children were very accepting of and interested in both languages, even for children whose parents did not speak Spanish. This anecdotal finding points out the need for further research on the attitudes of teachers, administrators, parents, and children in both all-English and dual-language public school settings. As it is not realistic to expect all public schools to become “immersion” schools, there may

instead be lessons public schools could learn from immersion schools in terms of promoting appreciation for linguistic diversity within and across children. Future qualitative and quantitative work could focus on attitudes teachers and children hold about language diversity in both English-only public schools and immersion schools, and what could be done to improve these attitudes in positive ways.

Findings from the current study also have important implications for policy. Several policies have been passed in recent years that reduce or eliminate bilingual education methods in public schools across several states, such as (but not limited to) Proposition 227 in California and Proposition 203 in Arizona (Crawford, 2000). The thinking process behind these policies is that DLLs will learn English quicker, and therefore perform better academically, if the focus in school is on English exclusively, as opposed to non-English languages. However, current study findings add to the growing body of literature that demonstrates that knowledge of two languages enhances EF, and EF is important for early school success (Blair, 2002). Policies that reduce or eliminate bilingual education should be reexamined, especially in terms of long-term academic and social outcomes for DLLs and balanced bilingual children.

Study findings also have implications regarding the importance of furthering theoretical understanding of the relationships between bilingualism, EF, and behavior problems during early and middle childhood, as EF is increasingly being viewed as important for school readiness (Blair, 2002). Indeed, many teachers report being just as, if not more, concerned with children's ability to pay attention, sit still, and follow directions upon entering kindergarten, as they are with children's knowledge of numbers

and letters (Lewit & Baker, 1995). Further, EF may provide an in-road for educators to intervene, both with bilingual children and children with externalizing behavior problems, to improve academic outcomes for these historically low-performing groups. The classroom environment is a landscape that provides children with opportunities to practice and refine EF skills that they already possess. Children's executive skills could be shaped and honed via adult modeling and classroom instructional strategies (Kaufman, 2010; Meltzer, 2010). Teachers could capitalize on the EF advantages displayed by bilingual children, as well as facilitate EF for children with behavior problems, to improve educational outcomes for these children. If the U.S. hopes to thrive and improve in its standing in the global economy, researchers, educators, policy-makers, and the general public should be concerned with the educational welfare of bilingual groups of students.



## REFERENCES

## REFERENCES

- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist 4-18 and 1991 profile*. Department of Psychiatry, University of Vermont, Burlington, VT.
- Achenbach, T. M. (2010). *Achenbach System of Empirically Based Assessment (ASEBA): ASEBA research updates from around the world*. Retrieved from: <http://www.aseba.org/research/research.html> on January 13, 2011.
- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*, 71-82.
- Anderson, P. J. (2008). Towards a developmental model of executive function. In V. Anderson, R. Jacobs, & P. J. Anderson (Eds.), *Executive functions and the frontal lobes: A lifespan perspective* (pp. 3-19). New York, NY: Taylor & Francis.
- Anderson, V., Anderson, P. J., Jacobs, R., & Smith, M. (2008). Development and assessment of executive function: From preschool to adolescence. In V. Anderson, R. Jacobs, & P. J. Anderson (Eds.), *Executive functions and the frontal lobes: A lifespan perspective* (pp. 123-154). Philadelphia, PA: Taylor & Francis.
- Atance, C. M., & Jackson, L. K. (2009). The development and coherence of future-oriented behaviors during the preschool years. *Journal of Experimental Child Psychology, 102*, 379-391.
- August, D., & Hakuta, K. (Eds.). (1998). *Educating language minority children*. Washington, DC: Committee on Developing a Research Agenda on the Education of Limited-English-Proficient and Bilingual Students, National Research Council and Institute of Medicine.
- Baddeley, A. (1986). *Working memory*. Oxford, England: Oxford University Press.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417-423.

- Baker, C. (2006). *Foundations of bilingual education and bilingualism* (4th ed.). Tonawanda, NY: Multilingual Matters Ltd.
- Baker, C., & Jones, S. P. (1998). *Encyclopedia of Bilingualism and Bilingual Education*. Philadelphia, PA: Multilingual Matters Ltd.
- Baker, S. A., Kovelman, I., Bialystok, E., & Petitto, L. A. (2003, November). Bilingual children's complex linguistic experience yields a cognitive advantage. *Published Abstracts of the Society for Neuroscience*, 1506.
- Barkley, R. A. (1997). *ADHD and the nature of self-control*. New York, NY: Guilford.
- Barkley, R. A., Grodzinsky, G., & DuPaul, G. (1992). Frontal lobe functions in attention deficit disorder with and without hyperactivity: A review and research report. *Journal of Abnormal Child Psychology*, 20, 163-188.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Bedore, L. M., Peña, E. D., Gillam, R. B., & Ho, T-H. (2010). Language sample measures and language ability in Spanish-English bilingual kindergarteners. *Journal of Communication Disorders*, 43, 498-510.
- Beitchman, J. H., Wilson, B., Brownlie, E. B., Waiters, H., Inglis, A., & Lance, W. (1996). Long-term consistency in speech/language profiles, 2: Behavioural, emotional and social outcomes. *Journal of the Academy of Child and Adolescent Psychiatry*, 35, 815-825.
- Benasich, A. A., Curtiss, S., & Tallal, P. (1993). Language, learning, and behavioral disturbances in childhood: A longitudinal perspective. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 585-594.
- Berk, L. E. (1992). Children's private speech: An overview of theory and the status of research. In R. M. Diaz & L. E. Berk (Eds.), *Private speech: From social interaction to self-regulation* (pp. 17-43). Hillsdale, NJ: Erlbaum.
- Berlin, L., & Bohlin, G. (2002). Response inhibition, hyperactivity, and conduct problems among preschool children. *Journal of Clinical Child and Adolescent Psychology*, 31, 242-251.
- Berlin, L., Bohlin, G., Nyberg, L., & Janols, L. O. (2004). How well do measures of inhibition and other executive functions discriminate between children with ADHD and controls? *Child Neuropsychology*, 10, 1-13.

- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development, 81*, 1641-1660.
- Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review, 29*, 180-200.
- Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70*, 636-644.
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, & cognition*. New York, NY: Cambridge University Press.
- Bialystok, E. (2009). Claiming evidence from non-evidence: A reply to Morton and Harper. *Developmental Science, 12*, 499-501.
- Bialystok, E. (2010). Global-local and trail-making tasks by monolingual and bilingual children: Beyond inhibition. *Developmental Psychology, 46*, 93-105.
- Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon Task. *Psychology and Aging, 19*, 290-303.
- Bialystok, E., & Feng, X. (2009). Language proficiency and executive control in proactive interference: Evidence from monolingual and bilingual children and adults. *Brain and Language, 109*, 93-100.
- Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics, 19*, 69-85.
- 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., & Shapero, D. (2005). Ambiguous benefits: The effect of bilingualism on reversing ambiguous figures. *Developmental Science, 8*, 595-604.
- Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition, 112*, 494-500.
- Bilancia, S. D., & Rescorla, L. (2010). Stability of behavioral and emotional problems over 6 years in children ages 4 to 5 or 6 to 7 at Time 1. *Journal of Emotional and Behavioral Disorders, 18*, 149-161.

- Bishop, D. V. M., Aamodt-Leeper, G., Creswell, C., McGurk, R., & Skuse, D. H. (2001). Individual differences in cognitive planning on the Tower of Hanoi task: Neuropsychological maturity or measurement error? *Journal of Child Psychology and Psychiatry* 42, 551-557.
- Bjorklund, D. F., & Harnishfeger, K. K. (1995). The evolution of inhibition mechanisms and their role in human cognition and behavior. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 141-173). San Diego, CA: Academic Press.
- Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, 57, 111-127.
- Borgmann, K. W. U., Risko, E. F., Stolz, J. A., & Besner, D. (2007). Simon says: Reliability and the role of working memory and attentional control in the Simon task. *Psychonomic Bulletin & Review*, 14, 313-319.
- Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching & Therapy*, 18, 1-22.
- Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology*, 26, 571-593.
- Brophy, M., Taylor, E., & Hughes, C. (2002). To go or not to go: Inhibitory control in hard-to-manage children. *Infant and Child Development*, 11, 125-140.
- Bub, K. L., McCartney, K., & Willett, J. B. (2007). Behavior problem trajectories and first-grade cognitive ability and achievement skills: A latent growth curve analysis. *Journal of Educational Psychology*, 99, 653-670.
- Bulotsky-Shearer, R. J., & Fantuzzo, J. W. (2011). Preschool behavior problems in classroom learning situations and literacy outcomes in kindergarten and first grade. *Early Childhood Research Quarterly*, 26, 61-73.
- Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. E. (2002). Immature frontal lobe contributions to cognitive control in children: Evidence from fMRI. *Neuron*, 33, 301-311.
- Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81-116). East Sussex, U.K.: Psychology Press.

- Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Developmental Psychology, 21*, 335-351.
- Cain, K., & Oakhill, J. (1996). The nature of the relationship between comprehension skill and the ability to tell a story. *British Journal of Developmental Psychology, 14*, 187-201.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.
- Campbell, S. B. (1994). Hard-to-manage preschool boys: Externalizing behavior, social competence, and family context at two-year follow-up. *Journal of Abnormal Child Psychology, 22*, 147-166.
- Campbell, S. B. (2002). *Behavior problems in preschool children: Clinical and developmental issues* (2nd ed.). New York, NY: Guilford.
- Campbell, S. B., Pierce, E. W., March, C. L., Ewing, L. J., & Szumowski, E. K. (1994). Hard-to-manage preschoolers: Symptomatic behavior across contexts and time. *Child Development, 65*, 836-851.
- Campbell, S. B., Shaw, D. S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology, 12*, 467-488.
- Carlson, S. M. (2003). Executive function in context: Development, measurement, theory, and experience. *Monographs of the Society for Research in Child Development, 68* (3, Serial No. 274), 138-151.
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595-616.
- Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*, 282-298.
- Carlson, S. M., & Moses, L. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032-1053.
- Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*, 299-319.

- Casey, B. J., Thomas, K. M., Welsh, T. F., Badgaiyan, R. D., Eccard, C. H., Jennings, J. R., & Crone, E. A. (2000). Dissociation of response conflict, attentional selection, and expectancy with functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences, 97*(15), 8728-8733.
- Connor, C. M., Ponitz, C. C., Phillips, B. M., Travis, Q. M., Glasney, S., & Morrison, F. J. (2010). First graders' literacy and self-regulation gains: The effect of individualizing student instruction. *Journal of School Psychology, 48*, 433-455.
- Couperus, J. W., & Nelson, C. A. (2006). Early brain development and plasticity. In K. McCartney & D. Phillips (Eds.), *The Blackwell handbook of early childhood development* (pp. 85–105). Oxford, UK: Blackwell.
- Crawford, J. (2000). English-only vs. English-only, a tale of two initiatives: California and Arizona. Retrieved on April 25, 2009, from <http://ourworld.compuserve.com/homepages/JWCrawford/203-227.htm>
- Crijnen, A. A. M, Achenbach, T. M., & Verhulst, F. C. (1997). Comparisons of problems reported by parents of children in 12 cultures: Total problems, externalizing, and internalizing. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 1269-1277.
- Crijnen, A. A. M, Achenbach, T. M., & Verhulst, F. C. (1999). Problems reported by parents of children in multiple cultures: The Child Behavior Checklist syndrome constructs. *The American Journal of Psychiatry, 156*, 569-574.
- Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Austin, TX: Pro-Ed.
- De Feyter, J. J., & Winsler, A. (2009). The early developmental competencies and school readiness of low-income immigrant children: Influences of generation, race/ethnicity, and national origins. *Early Childhood Research Quarterly, 24*, 411-431.
- de Villiers, P. (2004). Assessing pragmatic skills in elicited production. *Seminars in speech and Language, 25*, 57-71.
- de Villiers, P. A., & de Villiers, J. G. (2010). Assessment of language acquisition. *WIREs Cognitive Science, 1*, 230-244.
- Diamond, A. (1990). Developmental time course in human infants and infant monkeys, and the neural bases of inhibitory control in reaching. *Annals of the New York Academy of Sciences, 608*, 637-676.

- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 466-503). New York, NY: Oxford University Press.
- Diamond, A., & Kirkham, N. (2001, October). *Card sorting by children of 3 & 4 years and task switching by older children: Inhibition needed to overcome "attentional inertia."* Presented at the Cognitive Development Society Annual Meeting, Virginia Beach, VA.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test – III*. Circle Pines, MN: Dunn Educational Services, Inc.
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test – Revised*. Circle Pines, MN: Dunn Educational Services, Inc.
- Dunn, L. M., Padilla, E. R., Lugo, D. E., & Dunn, L. M. (1986). *Test de Vocabulario en Imagenes Peabody*. Circle Pines, MN: Dunn Educational Services, Inc.
- Durston, S., Davidson, M. C., Thomas, K. M., Worden, M. S., Tottenham, N., Martinez, A., ... Casey, B. J. (2003). Parametric manipulation of conflict and response competition using rapid mixed-trial event-related fMRI. *NeuroImage, 20*, 2135-2141.
- Eigsti, I. M., Zayas, V., Mischel, W., Shoda, Y., Ayduk, O., Dadlani, M. B... Casey, B. J. (2006). Predicting cognitive control from preschool to late adolescence and young adulthood. *Psychological Science, 17*, 478-484.
- Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals: Evidence from bimodal bilinguals. *Psychological Science, 19*, 1201-1206.
- Fagan, J., & Iglesias, A. (2000). The relationship between fathers' and children's communication skills and children's behavior problems: A study of Head Start children. *Early Education and Development, 11*, 307-320.
- Fahie, C. M., & Symons, D. K. (2003). Executive functioning and theory of mind in children clinically referred for attention and behavior problems. *Applied Developmental Psychology, 24*, 51-73.
- Fan, J., McCandliss, B. D., Sommer, T., Raz, M., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 340*, 340-347.



- Farmer, M. (1997). Exploring the links between communication skills and social competence. *Educational and Child Psychology, 14*, 38-44.
- Fernyhough, C., & Fradley, E. (2005). Private speech on an executive task: Relations with task difficulty and task performance. *Cognitive Development, 20*, 103-120.
- Festman, J., Rodriguez-Fornells, A., & Münte, T. F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions, 6*, Article 5.3
- Frye, D., Zelazo, P., & Palfai, T. (1995). Theory of mind and rule-base reasoning. *Cognitive Development, 10*, 483-527.
- Fujiki, M., Brinton, B., & Todd, C. M. (1996). Social skills of children with specific language impairment. *Language, Speech, and Hearing Services in the Schools, 27*, 195-202.
- Gandara, P., & Rumberger, R. W. (2009). Immigration, language, and education: How does language policy structure opportunity? *Teachers College Record, 111*(3), 750-782.
- Garcia, E. E., & Cuellar, D. C. (2006). Who are these linguistically and culturally diverse students? *Teachers College Record, 108*, 2220-2246.
- Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31-60.
- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 ½ - 7 years old on a Stroop-like day-night test. *Cognition, 53*, 129-153.
- Gioia, G. A., Espy K. A., & Isquith, P. K. 2002. *Behavior Rating Inventory of Executive Function, Preschool Version (BRIEF-P)*. Odessa, FL: Psychological Assessment Resources.
- Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior rating inventory of executive function*. Odessa, FL: Psychological Assessment Resources.
- Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C. ....Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences, 101*, 8174-8179.

- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*, 67-81.
- Greenhalgh, K. S., & Strong, C. J. (2001). Literate language features in spoken narratives of children with typical language and children with language impairments. *Language, Speech, and Hearing Services in Schools, 32*, 114–125.
- Guttentag, R. E., Haith, M. M., Goodman, G. S., & Hauch, J. (1984). Semantic processing of unattended words in bilinguals: A test of the input switch mechanism. *Journal of Verbal Learning and Verbal Behavior, 23*, 178–188.
- Hair, E., Halle, T., Terry-Humen, E., Lavelle, B., & Calkins, J. (2006). Children's school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first grade. *Early Childhood Research Quarterly, 21*, 431-454.
- Han, W-J. (2010). Bilingualism and socioemotional well-being. *Children and Youth Services Review, 32*, 720-731.
- Han, W-J., & Huang, C-C. (2010). The forgotten treasure: Bilingualism and Asian children's emotional and behavioral health. *American Journal of Public Health, 100*, 831-838.
- Harnishfeger, K. K. (1995). The development of cognitive inhibition: Theories, definitions, and research evidence. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 175-204). San Diego, CA: Academic Press.
- Hart, C. H., Olsen, S. E, Robinson, C., & Mandlco, B. L. (1997). The development of social and communicative competence in childhood: Review and model of personal, familial, and extrafamilial processes. *Communication Yearbook, 20*, 305-373.
- Hazen, N. L., & Black, B. (1989). Preschool peer communication skills: The role of social status and interactional context. *Child Development, 60*, 867-876.
- Heaton, R. K. (1981). *The Wisconsin Card Sorting Test Manual*. Odessa, FL: Psychological Assessment Resources.
- Heilmann, J., Iglesias, A., & Nockerts, A. (2008). Narrative transcription accuracy and reliability in two languages. *Topics in Language Disorders, 28*, 178-188.
- Hewitt, L. E., Hammer, C. S., Yont, K. M., & Tomblin, J. B. (2005). Language sampling for kindergarten children with and without SLI: Mean length of utterance, IPSYN, and NDW. *Journal of Communication Disorders, 38*, 197.

- Iaboni, F., Douglas, V. I., & Baker, A. G. (1995). Effects of reward and response costs on inhibition in ADHD children. *Journal of Abnormal Psychology, 104*, 232-240.
- Isquith, P. K., Crawford, J. S., Espy, K. A., & Goia, G. A. (2005). Assessment of executive functioning in preschool-aged children. *Mental Retardation and Developmental Disabilities Research Reviews, 11*, 209-215.
- Justice, L., M., Bowles, R., Pence, K., & Gosse, C. (2010). A scalable tool for assessing children's language abilities within a narrative protocol: The NAP (Narrative Assessment Protocol). *Early Childhood Research Quarterly, 25*, 218-234.
- Kalbfleisch, M. L. (2004). Functional neural anatomy of talent. *The Anatomical Record, Part B: New Anatomy, 277(1)*, 21-36.
- Kaufman, C. (2010). *Executive function in the classroom: Practical strategies for improving performance and enhancing skills for all students*. Baltimore, MD: Paul H. Brookes.
- Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence*. Hoboken, NJ: John Wiley & Sons.
- Kim, Y. K., Richard, E., & Winsler, A. (2012). Socio-emotional skills, behavior problems, and L1 competence predict the acquisition of English for English Language Learners in poverty. *Manuscript in preparation*.
- Klingner, J., & Artiles, A. J. (2006). English language learners struggling to learn to read: Emergent scholarship on linguistic differences and learning disabilities. *Journal of Learning Disabilities, 39*, 386-389.
- Klorman, R., Hazel-Fernandez, L. A., Shaywitz, S. E., Fletcher, J. M., Marchione, K. E., Holahan, J. M., Shaywitz, B. A. (1999). Executive functioning deficits in attention-deficit/hyperactivity disorder are independent of oppositional defiant or reading disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 38*, 1148-1155.
- Kopecky, H., Chang, H. T., Klorman, R., Thatcher, J. E., & Borgstedt, A. D. (2005). Performance and private speech of children with attention-deficit/hyperactivity disorder while taking the Tower of Hanoi Test: Effects of depth of search, diagnostic subtype, and methylphenidate. *Journal of Abnormal Child Psychology, 33*, 625-638.
- Kuntsi, J., Andreou, P., Ma, J., Borger, N. A., & van der Meere, J. J. (2005). Testing assumptions for endophenotype studies in ADHD: Reliability and validity of tasks in a general population sample. *BMC Psychiatry, 5*, Article 40.

- LaParo, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years. A meta-analytic review. *Review of Educational Research, 70*, 443–484.
- Lavigne, J. V., Arend, R., Rosenbaum, D., Binns, H., Christoffel, K. K., & Gibbons, R. D. (1998). Psychiatric disorders with onset in the preschool years: I. Stability of diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*, 1246-1254.
- Lehto, J., Juujarvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*, 59-80.
- Leung, P. W. L., & Connolly, K. J. (1996). Distractibility in hyperactive and conduct-disordered children. *Journal of Child Psychology and Psychiatry, 37*, 305-312.
- Levin, H. S., Culhane, K. A., Hartmann, J., Evankovich, K., Mattson, A. J., Harward, H... Fletcher, J. M. (1991). Developmental changes in performance on tests of purported frontal lobe functioning. *Developmental Neuropsychology, 7*, 377-395.
- Lewit, E. M., & Baker, L. S. (1995). School readiness. *The Future of Children, 5*, 128-139.
- Lidstone, J. S. M., Meins, E., & Fernyhough, C. (2010). The roles of private speech and inner speech in planning during middle childhood: Evidence from a dual task paradigm. *Journal of Experimental Child Psychology, 107*, 438-451.
- Lipsey, M. W., Wilson, S. J., & Farran, D. C. (2010). Measuring cognitive self regulation in relation to school readiness: Illustrative example. *Workshop: Executive function in preschool children: Current knowledge and research opportunities*. Vanderbilt University: Peabody Research Institute.
- Lu, C.-H., & Proctor, R. W. (2001). Influence of irrelevant information on human performance: Effects of S–R association strength and relative timing. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 54A*, 95–136.
- Luciana, M., & Nelson, C. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia, 36*, 273–293.

- Luciana, M., & Nelson, C. A. (2002). Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery: Performance in 4- to 12-year-old children. *Developmental Neuropsychology, 22*, 595-624.
- Luria, A. R. (1961). *The role of speech in the regulation of normal and abnormal behavior* (J. Tizard, Ed.). New York, NY: Pergamon Press.
- Mandell, D. S., Thompson, W. W., Weintraub, E. S., DeStefano, F., & Blank, M. B. (2005). Trends in diagnosis rates for autism and ADHD at hospital discharge in the context of other psychiatric diagnoses. *Psychiatric Services, 56*, 56-62.
- Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11*, 81-93.
- Matthew, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101*, 689-704.
- Mayer, M. (1974). *Frog goes to dinner*. New York: Dial Books.
- Mayer, M. (1969). *Frog, where are you?* New York: Dial Books.
- McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*, 947-959.
- McIntosh, K., Chard, D., Boland, J., & Horner, R. H. (2006). A demonstration of combined efforts in school-wide academic and behavioral systems and incidence of reading and behavior challenges in early elementary grades. *Journal of Positive Behavior Interventions, 8*, 146-154.
- Meltzer, L. (2010). Promoting executive function in the classroom. In K. R. Harris & S. Graham (Eds.), *What works for special needs learners* (pp. 3-27). New York, NY: Guilford.
- Miles, S. B., & Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Development, 77*, 103-117.
- Miller, J. F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research and Practice, 21*, 30-43.

- Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science, 244*, 933-938.
- 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.
- Moreland, A. D., & Dumas, J. E. (2008). Categorical and dimensional approaches to the measurement of disruptive behavior in the preschool years: A meta-analysis. *Clinical Psychology Review, 28*, 1059-1070.
- Moreno, S., Case, T., Schellenberg, E. G., & Bialystok, E. (2010, May). *Effect of bilingualism and musical training on executive control*. Poster presented at Development 2010, Carleton University, Ottawa, Ontario, CA.
- Moss, M., & Puma, M. (1995). *Prospects: The congressionally mandated study of educational opportunity and growth: First year report on language minority and limited English proficient students*. Washington, DC: U.S. Department of Education.
- Morton, J. B., & Harper, S. N. (2007). Bilinguals show an advantage in cognitive control—the question is why. *Developmental Science, 12*, 502-503.
- Mulas, F., Capilla, A., Fernandez, S., Etchepareborda, M. C., Campo, P., Maestu, F., ... Ortiz, T. (2006). Shifting-related brain magnetic activity in attention-deficit/hyperactivity disorder. *Biological Psychiatry, 59*, 373-379.
- Mullane, J. C., Corkum, P. V., Klein, R. M., & McLaughlin, E. (2009). Interference control in children with and without ADHD: A systematic review of flanker and Simon task performance. *Child Neuropsychology, 15*, 321-342.
- Muñoz, M. L., Gillam, R. B., Peña, E. D., & Gulley-Faehnle, A. (2003). Measures of language development in fictional narratives of Latino children. *Language, Speech, and Hearing Services in Schools, 34*, 332-342.
- Namazi, M., & Thordardottir, E. (2010). A working memory, not bilingual advantage, in controlled attention. *International Journal of Bilingual Education and Bilingualism, 13*, 597-616.
- National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs (NCELA). (2010). *The growing numbers of English learner students, 1997/98 – 2007/08*. Washington, DC: U.S. Department of Education.

- Nelson, C. A., & Bloom, F. E. (1997). Child development and neuroscience. *Child Development, 68*, 970-987.
- Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: View from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin, 126*, 220-246.
- Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders, 38*, 287-313.
- Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. Davidson, G. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4, pp. 1-18). New York, NY: Plenum Press.
- Okanda, M., Moriguchi, Y., & Itakura, S. (2010). Language and cognitive shifting: Evidence from young monolingual and bilingual children. *Psychological Reports, 107*, 68-78.
- Pennington, B. F. (1994). The working memory function of the prefrontal cortices: Implications for developmental and individual differences in cognition. In M. M. Haith, J. Benson, R. Roberts, & B. F. Pennington (Eds.), *The development of future oriented processes* (pp. 243-292). Chicago, IL: University of Chicago Press.
- Pennington, B. F., Grossier, D., & Welsh, M. C. (1993). Contrasting cognitive deficits in attention deficit disorder versus reading disability. *Developmental Psychology, 29*, 511-523.
- Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 37*, 51-87.
- Ponitz, C. E. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23*, 141-158.
- Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology, 45*, 605-619.
- Preciado, J. A., Horner, R. H., & Baker, S. K. (2009). Using a function-based approach to decrease problem behaviors and increase academic engagement for Latino English language learners. *The Journal of Special Education, 42*, 227-240.

- Raaijmakers, M. A. J., Smidts, D. P., Sergeant, J. A., Maassen, G. H., Posthumus, J. A., van Engeland, H., & Matthys, W. (2008). Executive functions in preschool children with aggressive behavior: Impairments in inhibitory control. *Journal of Abnormal Child Psychology, 36*, 1097-1107.
- Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 1-38). East Sussex, U.K.: Psychology Press.
- Rappport, M. D., Tucker, S. B., DuPaul, G. J., Merlo, M., & Stoner, G. (1986). Hyperactivity and frustration: The influence of control over and size of rewards in delaying gratification. *Journal of Abnormal Child Psychology, 14*, 181-204.
- Redmond, S. M. (2004). Conversational profiles of children with ADHD, SLI and typical development. *Clinical Linguistics and Phonetics, 18*, 107-125.
- Reitan, R. M. (1971). Trail Making Test results for normal and brain-damaged children. *Perceptual and Motor Skills, 33*, 575-581.
- Rice, M. L. (1993). "Don't talk to him. He's weird." A social consequence account of language and social interactions. In A. P. Kaiser & D. B. Gray (Eds.), *Enhancing children's communication: Research foundations for intervention* (pp. 139-158). Baltimore, MD: Paul H. Brookes.
- Rice, M. L., Sell, M., & Hgadley, P. A. (1991). Social interaction of speech- and language-impaired children. *Journal of Speech and Hearing Research, 34*, 1299-1307.
- Roger, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology, 124*, 207-231.
- Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology, 28*, 573-594.
- Rutter, M., Dunn, J., Plomin, R., Simonoff, E., Pickles, A., Maughan, B...Eaves, L. (1997). Integrating nature and nurture: Implications of person-environment correlations and interactions for developmental psychopathology. *Development and Psychopathology, 9*, 335-364.
- Sanford, A. (2006). *The effects of function-based literacy instruction on problem behavior and reading growth*. Unpublished doctoral dissertation, University of Oregon, Eugene.



- Seidman, L. J., Biederman, J., Faraone, S. V., Milberger, S., Seiverd, K., Benedict, K., & Ouellete, C. (1996). Toward defining a neuropsychology of ADHD: Performance of children and adolescents from a large clinically referred sample. *Journal of Consulting and Clinical Psychology, 65*, 150-160.
- Senn, T. E., Espy, K. A., & Kaufmann, P. M. (2004). Using path analysis to understand executive function organization in preschool children. *Developmental Neuropsychology, 26*, 445-464.
- Shallice, T. (1982). Specific impairments in planning. *Philosophical Transactions of the Royal Society of London B, 298*, 199-209.
- Shallice, T. & Burgess, P. W. (1996). The domain of supervisory processes and temporal organisation of behaviour. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 351*, 1405-1412.
- Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology, 27*, 960-974.
- Shin, H. B., & Bruno, R. (2003). Language use and English-speaking ability: 2000. *Census 2000 Brief*. U.S. Census Bureau.
- Shue, K., & Douglas, V. I. (1992). Attention deficit hyperactivity disorder and the frontal lobe syndrome. *Brain and Cognition, 20*, 104-124.
- Simmonds, D. J., Fotedar, S. G., Suskauer, S. J., Pekar, J. J., Denckla, M. B., & Mostofsky, S. H. (2007). Functional brain correlates of response time variability in children. *Neuropsychologia, 45*, 2147-2157.
- Simon, H. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7*, 268-288.
- Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology, 81*, 174-176.
- Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 31-86). Amsterdam: North-Holland.
- Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics, 30*, 315-337.

- Spira, E.G., & Fischel, J. E. (2005). The impact of preschool inattention, hyperactivity, and impulsivity on social and academic development: A review. *Journal of Child Psychology and Psychiatry, 46*, 755–774.
- Stroop, J. P. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.
- Stuss, D. T. (1992). Biological and psychological development of executive functions. *Brain and Cognition, 20*, 8-23.
- Stuss, D. T., Shallice, T., Alexander, M. P., & Picton, T. W. (1995). A multidisciplinary approach to anterior attentional functions. *Annals of the New York Academy of Sciences, 769*, 191-211.
- Trommer, B. L., Hoepfner, J., Lorber, R., & Armstrong, K. (1989). Pitfalls in the use of a continuous performance test as a diagnostic tool in attention deficit disorder. *Developmental and Behavioral Pediatrics, 9*, 339-346.
- Uccelli, P., & Páez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools, 38*, 225-236.
- U.S. Census Bureau. (No date). *Hispanic Americans by the Numbers*. Retrieved from: <http://www.infoplease.com/spot/hhmcensus1.html>.
- van Mourik, R., Papanikolaou, A., van Gellicum-Bijlhout, J., van Oostenbruggen, J., Veugelers, D., Post-Uiterweer, A....Oosterlaan, J. (2009). Interference control in children with attention deficit/hyperactivity disorder. *Journal of Abnormal Child Psychology, 37*, 293-303.
- Visser, S. N., Bitsko, R. H., Danielson, M. L., Perou, R., & Blumberg, S. J. (2010). Increasing prevalence of parent-reported attention-deficit/hyperactivity disorder among children – United States, 2003 and 2007. *MMWR Morbidity and Mortality Weekly Report, 59*, 1439-1443.
- Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: MIT Press.
- Wanless, S. B., McClelland, M. M., Acock, A. C., Ponitz, C. C., Son, S., Lan, X., ... Li, S. (2011). Measuring behavioral regulation in four societies. *Psychological Assessment, 23*, 364-378.

- Welsh, M., Pennington, B. F., & Grossier, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology, 7*, 131–149.
- Westby, C., Van Dongen, R., & Maggart, Z. (1989). Assessing narrative competence. *Seminars in Speech and Language, 10*, 63-75.
- Weyandt, L. L., & Willis, W. G. (1994). Executive functions in school-aged children: Potential efficacy of tasks in discriminating clinical groups. *Developmental Neuropsychology, 19*, 27-38.
- Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry, 57*, 1136-1346.
- Williams, B. R., Ponesse, J. S., Schachar, R. J., Logan, G. D., & Tannock, R. (1999). Development of inhibitory control across the life span. *Developmental Psychology, 35*, 205–213.
- Winsler, A. (1998). Parent-child interaction and private speech in boys with ADHD. *Applied Developmental Science, 2*, 17-39.
- Winsler, A. (2009). Still talking to ourselves after all these years: A review of current research on private speech. In A. Winsler, C. Fernyhough, & I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 3-41). New York, NY: Cambridge University Press.
- Winsler, A., Abar, B., Feder, M. A., Schunn, C. D., & Rubio, D. A. (2007). Private speech and executive functioning among high-functioning children with autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 37*, 1617-1635.
- Winsler, A., Diaz, R. M., Atencio, D. J., McCarthy, E. M., & Chabay, L. A. (2000). Verbal self-regulation over time in preschool children at risk for attention and behavior problems. *Journal of Child Psychology and Psychiatry, 41*, 875-886.
- Winsler, A., Fernyhough, C., & Montero, I. (2009). *Private speech, executive functioning, and the development of verbal self-regulation*. New York: Cambridge University Press.
- Zelazo, P. D. (2006). The dimensional change card sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*, 297-301.

## CURRICULUM VITAE

Lindsey A. Hutchison graduated from Middleburg High School, Middleburg, Florida, in 2002. She received her Bachelor of Science from the University of North Florida in 2006. She received her Master of Arts in Applied Developmental Psychology from George Mason University in 2009.