NARRATIVE COMPETENCE AND EXECUTIVE FUNCTIONING IN YOUNG CHILDREN WITH VARYING DEGREES OF BILINGUALISM

by

Danielle L. Mead
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: ___________________________________ Summer Semester 2015
George Mason University
Fairfax, VA

Narrative Competence and Executive Functioning in Young Children with Varying Degrees of Bilingualism

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

by

Danielle L. Mead
Master of Arts
George Mason University, 2013

Director: Adam Winsler, Professor
Department of Psychology

Summer Semester 2015
George Mason University
Fairfax, VA

Copyright 2015 Danielle L. Mead
All Rights Reserved

DEDICATION

This is dedicated to my parents, Colette and Joseph Mead, for their unfailing love and support, and for always believing in me.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Adam Winsler, for providing the guidance and support I needed to accomplish my goal, as well as the fellow students who have helped me in the coding of the narratives. I would like to recognize my family, friends, and loved ones for their emotional support throughout my pursuit of completing graduate school.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Behavior Rating Inventory of Executive Function .... BRIEF
Dimensional Change Card Sort .... DCCS
Comparative Fit Index .... CFI
Dual Language Learner...DLL
Executive Functioning ....EF
Exploratory Data Analysis.. EDA
Exploratory Factor Analysis ....EFA
Go/No-Go task ....GNG
Head-Toes-Knees-Shoulders task... HTKS
High Point Analysis ....HPA
Mean Length of Utterance ....MLU
Narrative Assessment Protocol ....NAP
Number of Different Words .... NDW
Number of Total Words .... NTW
Peabody Picture Vocabulary Test..PPVT
Root Mean Square Error of Approximation ....RMSEA
Story Grammar .... SG
Structural Equation Modeling... SEM
Test de Vocabulario en Imágenes Peabody ....TVIP
Tower of London ....TOL
United States .... U.S.

ABSTRACT

NARRATIVE COMPETENCE AND EXECUTIVE FUNCTIONING IN YOUNG CHILDREN WITH VARYING DEGREES OF BILINGUALISM

Danielle L. Mead, Ph.D.

George Mason University, 2015

Dissertation Director: Dr. Adam Winsler

Narratives are verbal accounts of an event—fictional or real—that follow a temporal sequence of clauses (Justice, Bowles, Pence, & Gosse, 2010; McCabe, 1997; Schick & Melzi, 2010). Narratives represent an interaction of linguistic, cognitive, and socio-cultural abilities, and are a culturally universal mode for expression and understanding experience. Researchers have emphasized the cognitive demands placed on an individual when creating a cohesive narrative, requiring the narrator to monitor the story's organization while presenting the causal and temporal sequence of events; however, there are few studies to explicitly test relations between narrative complexity and executive functioning. In bilingual samples, advantages in executive functioning have been established, and more recently, narratives are being used in lieu of or in addition to standardized assessments of expressive vocabulary because narratives are considered ecologically valid and culturally unbiased measures of language complexity (Bedore,

Peña, Gillam, & Ho, 2010; Fiestas & Peña, 2004). Thus, the current study examines the narrative structure and complexity of stories during a storytelling task and relates that to several direct measures of executive functioning. Additionally, language group (monolingual vs. bilingual), along with a continuous measure of degree of bilingualism, were tested as moderators to see if the relation between narrative complexity and executive functioning varies by the bilingualism status of the child. Narrative complexity was coded using several different measures to sensitively detect individual differences among the narratives produced in English and/or Spanish in (N = 79) 5- to 7-year-old children, consisting of English/Spanish bilinguals, dual language learners not yet fully proficient in a second language, and monolingual (English) children. Exploratory factor analyses (EFAs) yielded underlying constructs of narrative complexity and executive functioning; thus, a series of structural equation models was run to explore the relation between comprehensive measurements of narrative complexity and executive functioning. In addition, regression models were run to determine how degree of bilingualism may affect the relation between narrative complexity and executive functioning. Results indicated that age significantly predicted English narrative complexity (i.e., older children produced more complex narratives), and gender and receptive vocabulary, but not age, predicted Spanish narrative complexity (i.e., girls and children with stronger Spanish vocabulary produced more complex narratives). There was a positive association between overall narrative complexity and executive functioning, but this correlation was partially accounted for by controlling for age. Also, when narrative was separated into micro- (e.g., vocabulary and grammar) and macro-levels (e.g., organization and plot details) of analyses, only the macro-level was significantly correlated with executive functioning. When only examining the Spanish narratives of the bilingual children, the association between narrative complexity and executive functioning was not found. Several variables were tested to see if the relation between narrative complexity and executive functioning was different for children based on age, language group, or degree of bilingualism, but none of these variables altered the relation. This study contributes to the larger body of narrative research that has only highlighted the cognitive skills required to tell a coherent and well-organized narrative without ever directly relating narratives to executive functioning. Although direction of effect is still unclear, parents and teachers may consider providing children with numerous opportunities for practicing their narrative storytelling, as improvements in narrative may also lead to improvements in executive functioning. Future research can build off of the comprehensive models that were run in this study and apply them to other populations or in a longitudinal framework.

INTRODUCTION

Narratives are verbal accounts of an event—fictional or real—that follow a temporal sequence of clauses (Justice, Bowles, Pence, & Gosse, 2010; McCabe, 1997; Schick & Melzi, 2010). A personal narrative contains memory recall of a real event and is often referred to in the past tense, whereas a fictional narrative can be a composition of imagined events or a recall of an existing story (McCabe, Bliss, Barra, & Bennett, 2008). Historically, the use of personal narratives in the field of psychology was influenced by the psychoanalytic framework, relying on analyses for thematic content of patients' recounts of real or imagined events (McCabe, 1997; Nicolopoulou, 1997). Presently, narratives are considered to be ecologically valid and culturally unbiased language samples (Bedore, Peña, Gillam, & Ho, 2010; Fiestas & Peña, 2004). Narratives represent an interaction of linguistic, cognitive, and sociocultural abilities, and are a universal mode for expression and understanding of experience (Coelho, Liles, & Duffy, 1990; Silliman, Bahr, Brea, Hnath-Chisolm, & Mahecha, 2002). Gorman, Fiestas, Peña, and Clark (2011) also note that in addition to presenting a well-organized series of events, a "good" narrative also captures the attention and interest of the listener, a critical social component of the traditional use of oral narratives.

The Role of Children's Narrative Skills in Language and Cognitive Development

The basis for examining the role of children's narratives in language and cognitive development stems from research on the importance of decontextualized language skills for the development of literacy skills. Decontextualized language requires meaning to be conveyed through linguistic devices (e.g., syntax, vocabulary) about abstract objects or events that are removed from the immediate surroundings, as opposed to contextualized language, which relies heavily on non-linguistic means (e.g., gestures, facial expression), shared knowledge between the narrator and the listener, and contextual cues from the immediate environment (Cummins, 1991; Curenton & Justice, 2004). Decontextualized discourse is suggested to play a critical role in literacy skills (Dickinson & Snow, 1987; Westby, 1991), presumably through the use of literate language features (i.e., elaborated noun phrases, adverbs, conjunctions, and mental and linguistic verbs) aimed to convey explicit accounts of events and prevent ambiguity of meaning (Curenton & Justice, 2004). Terry, Mills, Bingham, Mansour, and Marencin (2013) state that because narrative discourse is decontextualized, storytelling can be considered as a cognitively complex skill, requiring the narrator to think beyond the immediate context to convey meaning. Although a wordless picture book may provide a small amount of contextual support, children must still rely on their language abilities to produce a cohesive and organized account of events.

The Role of Narratives in Monolingual and Bilingual Children

Research on children's narratives and their relation with language and literacy development originally began with English-speaking monolingual children, and revealed that narratives skills can predict academic achievement and school success (Dickinson & Snow, 1987; Westby, 1991). Literacy practices in the home environment, particularly through interactive narrative storytelling between parents and their preschoolers, have been shown to be predictive of literacy skills (Snow, 1983). In a randomized-control, 12-month narrative intervention study with monolingual preschoolers, mothers were asked to engage their preschoolers with narrative storytelling and ask open-ended questions to encourage conversation (Peterson, Jesso, & McCabe, 1999). Compared to the control group, the children in the intervention group demonstrated improved narrative ability in terms of linguistic complexity and quality as well as standardized receptive vocabulary both post-treatment and at a one-year follow-up (Peterson et al., 1999).

Narrative and Bilingualism

The evidence on the importance of narrative in monolinguals sparked an interest to explore narratives in other samples, such as Spanish-speakers (Gutierrez-Clellen & Heinrichs-Ramos, 1993; Gutierrez-Clellen & Hofstetter, 1994; Gutiérrez-Clellen & Iglesias, 1992), speakers of African American English (Burns, de Villiers, Pearson, & Champion, 2012; Schachter & Craig, 2013), Dual Language Learners (DLLs; Bailey, Moughamian, & Dingle, 2008; Bedore et al., 2011; Branum-Martin et al., 2009; Heilmann et al., 2008; Silliman et al., 2002; Solari, Zucker, Aghara, & Petscher, 2013), and balanced bilinguals (Cooperson, Bedore, & Peña, 2013; Dart, 1992; Fiestas & Peña, 2004; Goldstein, Bunta, Lange, Rodriguez, & Burrows, 2010; Iluz-Cohen & Walters, 2012; Jacobson & Walden, 2013; Miller et al., 2006; Uccelli & Páez, 2007; Uchikoshi, 2005). A practical reason for the interest in non-native English speakers is that the number of bilingual children in the United States is growing rapidly (Gándara & Rumberger, 2009), with a concomitant need to understand the challenges that these children face, especially in terms of language development. Bilingual children tend to perform lower than monolinguals on standardized assessments in receptive and expressive vocabulary (Pearson, Fernández, & Oller, 1993). This difference in performance may depend on the type of bilingual situation that the child is in (for example, recently immigrating to the United States) and whether the standardized language assessments use vocabulary (i.e., number of words a child knows within one language) or total conceptual vocabulary (i.e., giving credit for knowledge of the concept in either language) (Core, Hoff, Rumiche, & Señor, 2013). Assessments that only use vocabulary knowledge may not give enough credit for the conceptual knowledge that a child may know or be representative of a child's everyday language use. Thus, speech samples via narratives of bilingual children have been proposed as a better, more naturalistic method of language assessment (Fiestas & Peña, 2004). Narratives produced by the samples listed above typically yield greater variability in the number of total words and number of different words, compared to results of studies on monolingual children. For example, Solari et al. (2013) sought to test the psychometric properties of a modified, Spanish version of the published Narrative Assessment Protocol (NAP; Justice et al., 2010), and the results yielded larger standard deviations in Spanish than the original, English version.

To date, the role of narratives on literacy development in bilinguals has not received much attention, but there has been one study that aimed to explore the relation between oral language and reading proficiency in Spanish-speaking English Language Learners. Miller et al. (2006) collected narrative data on 1,531 children (in English and Spanish) from kindergarten through third grade, in addition to measures of story comprehension and reading efficiency. After controlling for age, oral narrative measures of production and quality significantly predicted reading outcomes, within each language, and interestingly, across languages. More specifically, cross-language effects were found for English oral language predicting Spanish reading efficiency and vice versa (Miller et al., 2006). The large sample in this study reflected a typical profile of bilingual children in the United States, and the results support that there is a relation between narrative performance and literacy skills. These results are excellent examples of how literacy is emerging in the field of narrative in bilingual children, especially because the associations between literacy and narrative had originally been found in monolinguals.

There are several studies that have examined narratives in bilingual children. Early studies by Gutierrez-Clellen and colleagues focused specifically on the narratives produced by Spanish-speakers (Gutierrez-Clellen & Heinrichs-Ramos, 1993; Gutierrez-Clellen & Hofstetter, 1994; Gutiérrez-Clellen & Iglesias, 1992), but many studies within the past 20 years have compared narratives produced in both languages of the child, which provide better insight as to the differences and similarities in narrative production of bilingual children within and across languages.

Grammatical errors. Narratives are sensitive at capturing a wide variety of key language skills that may not be measured using standardized assessments of receptive and expressive language, such as grammatical errors. In dual language learners—who sequentially learn a second language after their dominant language—there is evidence that grammatical errors in narratives are produced differentially across languages, with proportionately more grammatical errors produced in their non-native language (Bailey et al., 2008; Bedore et al., 2010; Cooperson et al., 2013). For balanced bilinguals, however, children have been found to produce the same amount of grammatical errors across languages (Fiestas & Peña, 2004; Gutiérrez-Clellen & Kreiter, 2003).

Narrative quality. Measures of narrative quality, such as the use of story elements (e.g., orienting the listener to the setting, stating characters' attempts to solve a problem) and organization (e.g., introduction, climax, resolution), have been compared across the two languages of bilinguals. In a study of 20 Spanish-dominant English Language Learners (ELLs), Bailey et al. (2008) compared the narrative structure between English and Spanish. The highest level of narrative quality found in Spanish followed a chronological order of events, without ever reaching a highpoint in the story. Within the narratives that were classified as containing a "classic" narrative structure, there were three times more classically-constructed English stories as there were Spanish ones, which may reflect a cultural difference in the preferred style of telling a story. The results of this study indicate that ELLs were able to produce well-constructed narratives in their non-dominant language of English. For bilinguals, however, language differences in narrative quality have not been found (Fiestas & Peña, 2004; Silliman et al., 2002), such that stories in English and Spanish were similar in terms of story grammar complexity and causal complexity.

Within- and cross-language abilities. Among bilingual children, associations between narrative and standardized language assessments have been established, with a particular emphasis on within-language and cross-language effects. A longitudinal study of narrative development in young bilingual children found that English standardized vocabulary scores in kindergarten significantly predicted narrative quality in first grade for both English and Spanish, but Spanish standardized vocabulary scores only predicted Spanish narrative quality (Uccelli & Páez, 2007). Branum-Martin et al. (2009) conducted a large-scale study of 1,300 children nested within 247 kindergarten and first-grade classrooms in an attempt to explore the relations between Spanish and English vocabulary. Vocabulary was measured by the picture-naming task in the Woodcock Language Proficiency Battery-Revised (Woodcook, 1991; Woodcock & Muñoz-Sandoval, 1995), as well as the number of different words produced during a narrative elicitation from a wordless picture book in English and Spanish. At the classroom- and student-level, the results yielded moderate to strong positive correlations between vocabulary from the picture-naming and narrative tasks within each language, but only across languages for the narrative task.

In addition to their empirical study, Branum-Martin and colleagues (2009) also conducted a meta-analysis to explore previous findings of correlations between Spanish and English vocabulary measures by separating the type of vocabulary measure used. Their findings clearly indicated that correlations of English and Spanish vocabulary obtained via narrative measures are consistently positive, while correlations between vocabulary scores obtained via other measures (e.g., standardized assessments of receptive vocabulary, expressive vocabulary) show great amounts of variability across studies, suggesting that narratives may measure a fundamentally different aspect of vocabulary as a construct (Branum-Martin et al., 2009). It is possible that the difference lies in the everyday nature of the language used in narrative tasks, in contrast to the academic nature of the vocabulary used in standardized assessments of expressive and receptive language. Therefore, as Fiestas and Peña (2004) argue, narratives may be an ecologically valid measure of language proficiency.

Development of Narrative Skills

There is one study that examined longitudinal change in narrative quality from kindergarten to first grade within a sample of 24 Spanish/English bilingual children in poverty (Uccelli & Páez, 2007). It was evident that there was a wide range of variability in terms of English and Spanish standardized vocabulary at both time-points, and narrative quality improved for both English and Spanish. As previous studies have suggested, the number of different words was sensitive to developmental change, as opposed to total number of words, but this difference was only found in English. Interestingly, there were cross-language effects, such that Spanish story structure in kindergarten significantly predicted English narrative quality in first grade, after controlling for English vocabulary (Uccelli & Páez, 2007).

The remaining studies of age-related differences in narratives have used cross-sectional methods, comparing different age groups on a variety of narrative production and quality measures. Curenton and Justice (2004) collected narratives of 3-, 4-, and 5-year-olds, and found more mental and linguistic verbs in the 5-year-olds compared to the

3-year-olds, suggesting that cognitive-based mental state terms become more prevalent over the preschool period. Therefore, if one is interested in the use of mental-state terms in narratives, it would be appropriate to include older, 5- to 7-year-old children. Gutierrez-Clellen and Heinrichs-Ramos (1993) examined referential cohesion (e.g., introduction and references to characters and places in a story) in 4-, 6-, and 8-year-old children, and found that older children were more capable of incorporating descriptive referential information into their narratives. Gutierrez-Clellen and Hofstetter (1994) also found that older children (6- and 8-year-olds) were more capable of differentiating characters in a story and elaborating on events and actions compared to younger children (5-year-olds). In addition, Muñoz, Gillam, Peña, and Gulley-Faehnle (2003) found that among 4- and 5-year-old English-speaking Latino children, older children produced more sophisticated narratives, while using more clauses and fewer syntactic errors. The results of these studies suggest there may be developmental changes in terms of the elaboration and complexity of narrative skills throughout the early school ages, and that global measures of narrative quality may be an appropriate focus for exploration, as opposed to the simpler, linguistic changes based on grammar and vocabulary. One must also consider that early school age is a critically important time in literacy development, and since narrative development has been related to literacy skills, researchers should draw attention to 5- to 7-year-olds.

Children's Narrative Skills and Executive Functioning

Executive functioning is an umbrella term that refers to the conscious cognitive processes needed to solve problems, such as planning, monitoring progress, inhibiting automatic responses, directing attention, and cognitive flexibility (Burgess, 1997; Diamond, 2012). The role of narrative in executive functioning has not been as explicitly studied in typically developing children, but researchers have discussed the importance of using a cognitive approach to studying narrative. For example, McCabe (1997) suggests analyzing children's narratives as they pertain to cognitive constructs, such as identifying the characters' goals and the problem-solving skills that the characters must use in order to obtain or resolve their goals. In order to fully comprehend a story, a child must create a mental representation of the story and then organize it (Skarakis-Doyle & Dempsey, 2008). However, there are a handful of studies with adults (Cannizzaro & Coelho, 2013), children with ADHD (Ygual Fernández, Roselló Miranda, & Miranda Casas, 2010), children with pragmatic language impairment (Ketelaars, Jansonius, Cuperus, & Verhoeven, 2012), and children who have experienced a traumatic brain injury (Brookshire, Chapman, Song, & Levin, 2000; Chapman et al., 1992) that have examined the association between narrative and cognitive assessments relating to executive functioning. In adults, the story grammar structure of narratives has been found to be correlated with measures of executive functioning, for both linguistic (e.g., verbal fluency, Stroop) and non-linguistic (e.g., Tower of London, motor speed) tasks (Cannizzaro & Coelho, 2013). In children with ADHD, unique features of executive functions have been found that predict different aspects of narrative quality; more specifically, planning predicted coherence, attention, and monitoring, while interference control predicted comprehension in narrative retellings of a story (Ygual Fernández et al., 2010).

Because traumatic brain injury (specifically, closed head injury; CHI) patients have been found to display language and cognitive deficits after their injury, researchers began to wonder how their narrative discourse would appear one, or up to three years after their injury. It has been suggested that because narrative discourse requires an individual to rely on language and cognitive abilities to plan, organize, and generate a cohesive, goal-directed narrative, patients with traumatic brain injury would perform poorly compared to healthy controls. Chapman and colleagues (1992) compared narrative quality between 20 CHI children and adolescents and 20 controls (i.e., matched on age, gender, and SES), and also compared performance in verbal memory. The authors analyzed the CHI children by level of severity, which resulted in a mild/moderate group and a severe group. Interestingly, all CHI children and controls did not differ on a measure of verbal memory and the flow of information in narratives, such that all participants were similar in terms of number of hesitations, repetitions, and revisions. CHI children and controls did, however, differ in narrative organization structure, and the severe CHI group produced significantly fewer number of words, number of clauses, and episode features (e.g., describing the setting, action sequences) compared to the control group. These results suggest that disruption in story structure may reflect impairment in the mental schemas that are used to create a cohesive story (Chapman et al., 1992).

Although the study described above used verbal memory as an indicator of cognitive functioning, there is one study that measured a wide array of executive functioning in addition to collecting narrative discourse in CHI patients. Brookshire et al. (2000) recruited 23 mild CHI and 68 severe CHI children and adolescents, ranging in age from 8 to 17, roughly three years after their brain injury. The subjects were asked to retell a narrative and answer comprehension questions based on the story, and then completed a variety of executive functioning tasks (e.g., Wisconsin Card Sort Test (WCST), Word Fluency, Divided Attention, Tower of London), processing speed (e.g., go/no-go tasks, WISC-R coding), and word recall. Differences were found between mild and severe CHI groups on word fluency, divided attention, and word recall, with the mild CHI patients outperforming the severe patients. Spearman rank correlations were run between narrative discourse structures and the cognitive assessments, and revealed that there were significant correlations for the WCST, word fluency, and Tower of London. The authors describe that story generation requires identifying a goal, planning the attainment of the goal, and evaluating the success or failure of the plan, all of which are important in mental organizational schemata and are found to be impaired in severe CHI patients (Brookeshire et al., 2000).

Bilingualism and Executive Functioning

A large portion of children in the world are exposed to more than one language (Bialystok, 2005), and bilingualism plays an important role in the development of language and cognition. Simultaneous bilinguals are exposed to more than one language from infancy, and researchers suggest that the ability to differentiate phonemes from different languages is strongest before 12 months of age (Bialystok, 2001; Hohle, 2009). Simultaneous bilinguals are likely to become balanced bilinguals, demonstrating similar levels of language abilities across the two languages (Bialystok, 2001), but research on balanced bilinguals in preschool and early school age children has predominantly been conducted in countries outside the United States. Therefore, there is a need to explore balanced bilingualism in American children. Sequential bilinguals are exposed to one language and then introduced to a second language (Bialystok, 2001). Dual Language Learners (DLLs) are individuals who do not speak the majority language as their primary language (for example, in the United State, those who do not speak English as their primary language at home) until a point in age when exposure to a second language begins. As the number of dual language learners (DLLs) in the United States continues to grow rapidly (Gándara & Rumberger, 2009), the educational needs of this special population are becoming more relevant, especially because DLLs typically fall behind academically compared to their monolingual peers (Espinosa, 2010; Rumberger & Tran, 2009). Thus, it is important to focus on children who may be exposed to more than one language, whether they are balanced bilinguals or DLLs, at the time in which they are entering elementary school.

Bialystok’s work on bilingualism has been notably influential (Bialystok, 2001), especially in regard to executive functioning. Among the various manifestations of executive functioning, balanced bilinguals are said to have an advantage over monolinguals in selective attention and inhibition (Bialystok, 2001). Theoretically, balanced bilinguals need to continuously know when and how to switch from one language to another. This switching between languages results in habitually resolving a conflict of two competing language systems, a process that takes place in frontal lobe executive processes (Bialystok, Craik, & Luk, 2008). In other words, bilinguals are always inhibiting one language when speaking the other language, and this practice in inhibition is theorized to improve performance on tasks that require selective attention and interference control (Bialystok, 2011). Assessments of interference control typically involve presenting problems with distracting and/or irrelevant factors that must be ignored in order to solve the problem or complete the task successfully. Balanced bilinguals have been shown to have an advantage in these control tasks compared to monolinguals (Bialystok & Majumder, 1998; Bialystok & Martin, 2004; Martin-Rhee & Bialystok, 2008; Poarch & Van Hell, 2012; Yang, Yang, & Lust, 2011).

To demonstrate the bilingual advantage, researchers commonly use the Simon task, which requires both selective attention and interference control. In the Simon task, there is a perceptual conflict that must be resolved, such that the individual must push a button on one side of a keyboard when one stimulus is presented on a screen, and another button must be pressed when a second stimulus appears on the screen. When the stimulus appears on the same side of the screen as the appropriate button, it is considered a congruent trial, whereas an incongruent trial is when a stimulus appears on the opposite side of the screen from where the designated button is on the keyboard. Theoretically, bilinguals should perform more quickly and successfully on incongruent trials, compared to congruent trials, because these trials demonstrate a conflict that requires selective attention, a skill that balanced bilinguals practice from an early age (Bialystok, 2011; Hutchison, 2012; Martin-Rhee & Bialystok, 2008). Empirically, this advantage on incongruent trials has been found (Martin-Rhee & Bialystok, 2008; Poarch & Van Hell, 2012), although one study found an advantage in both incongruent and congruent trials (Bialystok, 2011).

Other areas of executive functioning are not necessarily associated with a bilingual advantage, although there has been some empirical support for inhibitory control and working memory tasks. In simple tasks of inhibitory control, individuals are required to suppress a prepotent, dominant response, such as go/no-go tasks (e.g., responding when one visual stimulus appears on a screen and not responding when a different stimulus appears) and Stroop tasks (Bialystok & Viswanathan, 2009; Carlson & Meltzoff, 2008; Iluz-Cohen & Armon-Lotem, 2013; Poulin-Dubois, Blaye, Coutya, & Bialystok, 2011). Working memory has been suggested to be a component of executive functioning (Baddeley, 1986, 2000), but it is difficult to differentiate working memory from higher-order executive functioning components that require working memory, such as planning and shifting. Therefore, it is not surprising that when working memory is measured in simple memory span tests, there are no differences among bilingual and monolinguals (Gutiérrez-Clellen, Calderón, & Weismer, 2004; Martin-Rhee & Bialystok, 2008). In adults, there have been similar findings for forward digit span, although an advantage was found in a backward digit span task, which is a more complex working memory task that requires one to mentally hold information and perform a reverse operation (Bialystok et al., 2008).

It is possible, however, that positive associations between bilingualism and executive functioning may exist only for balanced bilinguals, that is, individuals who have comparable language abilities in two languages. For DLLs, one language is thought of as dominant (Baker, 2006), and research has not found the advantages in executive functioning for DLLs that have been found in balanced bilinguals (Baker, 2006;

Bialystok & Majumder, 1998; Carlson & Meltzoff, 2008; Luciana & Nelson, 1998). It is possible that the simultaneous exposure to two languages from birth allows a bilingual child to practice managing their two language systems, whereas learning languages sequentially does not provide the same degree of practice opportunities, in addition to less overall exposure to their second language. Hutchison (2012) compared direct measures of executive functioning in balanced bilinguals, dual language learners (DLLs), and monolinguals. Degree of bilingualism was not related to simpler inhibition tasks or parent report of executive functioning, but there was a balanced bilingual advantage in complex inhibitory control (i.e., the Simon task), and DLLs demonstrated better performance than monolinguals in tasks measuring interference control and cognitive flexibility. Hutchison also treated degree of bilingualism as a continuous variable based on creating standardized *z*-scores of the composite of differences between English and Spanish skills on a variety of language measures. This study provides promising results in the exploration of treating degree of bilingualism as a continuous variable, which may be more sensitive to detect differences among children with exposure to two languages than grouping all bilingual children together.

Development of Executive Functioning

Research on investigating executive functioning and its role in child development has focused on children as young as age two (Carlson & Meltzoff, 2008) through the preschool years. There is less of an emphasis on school-age children (Hutchison, 2012), and some measures of executive functioning have been designed primarily for preschoolers. However, there have been cross-sectional studies of executive functioning from preschool through adolescence, resulting in a sharp increase of performance on tasks in 4- and 5-year-olds, followed by a steady increase through 8-year-olds, and plateauing off until age 12 (Zelazo et al., 2013; Zelazo & Carlson, 2012). Although these data do not provide information on individual growth in executive functions through adolescence, they do shed light on what may be going on during the early school years. Therefore, there should be more attention drawn to ages 5 through 7, especially in terms of the validity of executive functioning measures used during this age range.

Methods of Obtaining Children's Narratives

In order to measure children's narrative skills, one must decide to either use spontaneous production of stories, or find another way of measuring narrative use and competence in children without necessarily asking them to produce a narrative. Indirect measures of language, such as asking parents global questions on an early literacy questionnaire, are considered easier than obtaining, transcribing, and coding a language sample; however, direct measures of language provide a more representative snapshot of a child's language (Goldstein et al., 2010). In terms of measuring narrative competence, some researchers have developed protocols to ask children direct questions about story elements, thus providing information on their narrative comprehension. Researchers have compared the length, grammaticality, and coherence of responses to narrative retelling of the story, and found that responses to direct questions were longer in length, richer in vocabulary, and more coherent than the narrative retell task (Gazella & Stockman, 2003; de Quirós, Lara-Alecio, Tong, & Irby, 2012; Silva, Strasser, & Cain, 2014).

Traditionally, researchers have measured children's narrative skills through direct assessment by obtaining speech samples. Direct narrative assessment allows children to produce a spontaneous speech sample that is then recorded and analyzed for varying degrees of complexity, whether it be linguistic complexity (e.g., total number of words, grammar errors) or narrative quality (e.g., the cohesion and organization of the story itself). Over the past 25 years, researchers have used several different elicitation procedures to obtain children's narrative skills. One procedure simply requires the child to tell an open-ended, personal story (Bailey et al., 2008; McCabe et al., 2008) or to complete a story after being given a story stem (McCabe, 1997) or a picture stimulus (Coelho et al., 1990; Justice et al., 2006; Uccelli & Páez, 2007). The most commonly used procedure uses a wordless picture book, and children are typically asked to look through the book (or first few pages) and then generate a story based on the pictures. Once a story is produced, researchers typically transcribe and code for the constructs of interest, which will be discussed further below.

Components of Interest in Narrative Speech Samples

When designing a study of narrative speech samples, researchers decide on which aspect of the narratives to focus on: the micro- or macro-structure level. If someone were interested in productivity, accuracy, grammaticality, syntax, and other linguistic forms occurring in speech samples, then a micro-structure framework would be appropriate (Bailey et al., 2008; Heilman, Miller, Nockerts, & Dunway, 2010). Studies of micro-structure tend to report variables such as the number of total words, number of different words, total number of clauses/utterances, mean length of clause/utterance (in words or morphemes), percentages of grammatical errors, and for bilinguals, code-switched utterances (i.e., containing one or more words of another language embedded within the same phrase or sentence). In general, these studies aim to use narrative discourse as a non-standardized method of capturing general language abilities. If, on the other hand, someone were interested in organization, thematic coherence, concepts, complexity, and anything else beyond the utterance level produced in narratives, then a focus on macro-structure would be best (Burns et al., 2012; Heilmann et al., 2010). Studies of macro-structure tend to report total number of episodes, mean length of complete episodes, judgments of ordinal levels of organization, and other meta-level variables within narrative storytellings (Heilmann et al., 2010; Lofranco, Peña, & Bedore, 2006; McCabe et al., 2008; Miller et al., 2006; Muñoz et al., 2003; Uccelli & Páez, 2007; Uchikoshi, 2005).

Although most studies use only one main structure level on which to focus, there are a handful of studies that have examined narratives at both levels (Bailey et al., 2008; Fiestas & Peña, 2004; Terry et al., 2013). In a study by Bailey and colleagues (2008), Spanish-dominant DLLs were found to produce fewer grammatical errors in Spanish narratives compared to English narratives, but their English narratives were more classically organized to build up to a climax followed by a resolution compared to their Spanish narratives. In a longitudinal study of preschool children, Terry et al. (2013) found that micro-level measures were sensitive enough to detect change over the course of the preschool year, but only one macro-level (i.e., High Point Analysis) was able to detect developmental change.

Rationale for the Current Study

Within narrative studies of monolinguals, researchers have emphasized that there are cognitive demands placed on an individual when creating a cohesive narrative, requiring the narrator to monitor the story's organization while presenting the causal and temporal sequence of events (Bailey et al., 2008; Coelho et al., 1990). Other researchers have focused on the goal-directed nature of narratives, such that children must identify the character's goal, describe how the character must problem-solve to attain the goal, and understand how obstacles along the way contribute to how the goal is to be accomplished (McCabe, 1997; Skarakis-Doyle & Dempsey, 2008). Children must rely on mental schemas of how stories should be structured and then apply their knowledge to the story presented to them (McCabe, 1997). These cognitive demands appear to be different manifestations of executive functioning, and yet narrative researchers have not explicitly referred to executive functioning in their studies. Fiestas and Peña (2004), who studied narratives of bilingual children, have also emphasized the cognitive abilities required for narrative discourse, including the need to plan, organize, and understand the meaning underlying the events, while simultaneously monitoring how the story is being presented in a cohesive manner and using complex sentence structure to demonstrate the causal sequence of events. Therefore, it is important to code narrative speech samples for aspects beyond the number of words and total length of a story in order to capture the important cognitive abilities that go into storytelling. A combination of coding for microstructural and macrostructural elements provides a comprehensive and thorough examination of narrative structure and complexity, allowing for the ability to test the sensitivity of each level of measurement, in addition to statistically testing for the existence of an underlying complexity construct among the various complexity scores. Thus, by using a combination of different levels of narrative elements, the data in the current study were rich and reflective of the organizational and planning skills that have been associated with cognitive features.

Heilman and colleagues (2010) have suggested that the organization skills children use to produce narratives stem from general cognitive abilities, such as executive functioning. The rationale for research on narrative discourse in children with traumatic brain injury is that narratives are hypothesized to require language, memory, and executive functions, such as planning, creating, and applying mental representation, in order to contribute to the narrative's macro-structure (Brookshire et al., 2000). As discussed above, bilinguals have been shown to have an advantage over monolinguals in a variety of executive functions, in particular those that require planning and monitoring. Therefore, if narratives are related to executive functioning, it would seem appropriate to presume that bilinguals would demonstrate better performance on the macro-structure (i.e., organization and cohesion) of narrative discourse compared to monolinguals. This hypothesis remains to be answered, and it may be possible that the bilingual advantage may only reside in balanced bilinguals because sequential bilinguals (or dual language learners) have not consistently shown increased performance on executive functioning tasks compared to monolinguals (Carlson & & Meltzoff, 2008). Therefore, further exploration in the effect of degree of bilingualism on executive functioning should be considered.

Research Questions

The purpose of this study is to analyze the narrative structure and complexity of stories during a storytelling task and relate that to several direct measures of executive functioning (e.g., cognitive flexibility, interference control, inhibition of a prepotent response, planning, inhibitory control). Additional analyses include testing if there are language differences (English-Spanish) in narrative storytelling abilities, testing the potential moderating effect of bilingual status to determine if the relation between narrative skills and executive functioning differs for English/Spanish bilinguals, dual language learners not yet fully proficient in L2, and monolingual (English) children, and among the bilingual children, determining if the relation is different for English or Spanish. Several measures of narrative complexity—at the microstructure and macrostructure level—were coded to sensitively detect individual differences among the narratives produced in English and/or Spanish in 5- to 7-year-old children. The goal of this study was to answer the following research questions:

(1) What are the predictors of narrative complexity (e.g., age, gender, receptive language skills)? Based on the findings of previous studies, it was hypothesized that older children would produce more elaborate and complex narratives than younger children, and that receptive language skills would be positively correlated with narrative complexity in both English and Spanish.

(2) In bilingual children, are there language differences (between English and Spanish) in narrative complexity? In balanced bilingual children, it was hypothesized that their performance would be similar in both English and Spanish. Since DLLs have stronger language skills in their dominant language compared to a non-dominant

language, it was hypothesized that narratives will be more complex in their dominant language (mostly English) compared to narrative produced in their non-dominant language (Spanish).

(3) How does narrative complexity relate to executive functioning across all children? The limited amount of research on narrative complexity and executive functioning suggests that there is a positive association between narrative complexity and executive functioning, such that children who produce more complex and organized narratives have better performance on executive functioning tasks compared to children who have poorly organized narratives.

(4) How does the relationship between narrative complexity and executive functioning vary by degree of bilingualism and/or age? Since bilingual children have been shown to have advantages among a variety of executive functioning skills, it was hypothesized that bilingual status will moderate the relation between narrative complexity and executive functioning. More specifically, it was hypothesized that this relation will be stronger for balanced bilingual children compared to monolinguals or DLLs.

(5) In DLL and bilingual children, is the relation between narrative complexity and executive functioning the same children whose parent-report of first language was English, Spanish, or both English and Spanish?

(6) Does the relation between narrative complexity and executive functioning depend on the level of analysis of narrative complexity (i.e., micro- vs. macro-level measures)? Micro-level measures provide information on the amount of language used to

tell a story, whereas macro-level measures include codes for planning and goal-directed attempts and consequences, as well as a logical, organized, causal sequence of events. Therefore, it was hypothesized that the macro-level measures would be more strongly associated with executive functioning compared to micro-level measures for all children.

METHOD

Participants

Seventy-nine children, between the ages of 5 and 7 years, were recruited to participate in a larger study of bilingualism, executive functioning, and behavior problems (for more details, see Hutchison, 2012). Thirty-three participants (41.77%) were monolingual English speakers and the remaining 46 (58.23%) were considered bilingual Spanish-English speakers, including both balanced bilinguals (*n* = 17) and Dual Language Learners (DLLs; *n* = 29) (see below for a detailed description of the language group criteria).

For the entire sample, there was a relatively even number of males (*n* = 39) and females (*n* = 40), the mean age was 74.67 months (*SD* = 9.96), the average annual family income was high ($60,000 – $80,000), and the majority of the participants were White (*n* = 43, 55.8%) or Hispanic (*n* = 40; 37.7%). Among the three language groups, the majority of the monolinguals were White (84.8%) and the remaining monolinguals were classified as Hispanic (3.0%) or Other (12.1%); the balanced bilinguals were three-quarters Hispanic and one-quarter White; and just over half of the DLLs were Hispanic (57.1%) and the remaining DLLs were classified as either White (39.3%) or Other (3.6%).

Language Group Criteria

Language group was determined by six criteria from both parent-report and direct assessment: (1) minimum exposure of 3 years to each language, (2) mean parent-report of proficiency score of 3 or higher for each language (1 = Poor, 5 = Excellent), (3) mean parent-report of language use score between 2.5-3.5, (4) English receptive language (i.e., PPVT) age equivalency score within the confidence interval according to chronological age of the child, (5) Spanish receptive language (i.e., TVIP) age equivalency score within the confidence interval according to chronological age of the child, and (6) a difference of less than 50 in the number of unique words produced in English and Spanish. A child must have met five of the six criteria to be considered a “balanced” bilingual. To be considered a “Dual Language Learner,” a child must have met two of the criteria, and have a minimum of 6 months of substantial exposure to a second language. Within the DLL group, based on the combination of the use and input of each language based on parent-report and the performance on the direct language assessments, the majority (n = 26, 89.66%) were classified as English dominant, and the remaining three were classified as either Spanish dominant (n = 2) or could not be classified due to inconsistent language data (n = 1). The English-dominant nature of the DLL group was largely due to recruitment limitations, such that many of these children were English-speaking, Anglo children attending Spanish immersion schools. For all children, parents completed a language background questionnaire, which included several items on exposure to and use of English and Spanish. These language items were used to determine if the parent reported that the child’s first language was English, Spanish, or that the child’s “first language” was both English and Spanish (i.e., 1 = English, 2 = Spanish, 3 = both English and Spanish).

Degree of Bilingualism

A primary goal of the larger study (Hutchison, 2012) was to calculate a continuous measure of the degree of bilingualism. This variable was calculated by calculating the difference score on each of the following language measures: PPVT/TVIP, number of unique words in English/Spanish during the story, parent-report of English/Spanish proficiency, and parent-report of English/Spanish use; taking the absolute value of each difference score; standardizing the absolute values of difference scores; and then averaging the standardized difference scores. For this variable, larger (more positive) scores reflected less balance and smaller (more negative) scores reflected more balance in degree of bilingualism.

Procedure

After obtaining informed parental consent and child assent, a variety of language assessments were administered, including measures of receptive English skills (Peabody Picture Vocabulary Test; PPVT; Dunn & Dunn, 1997) and receptive Spanish skills (Test de Vocabulario en Imágenes Peabody; TVIP; Dunn, Padilla, Lugo, & Dunn, 1986) for all children, regardless of language group to ensure that monolingual English speakers had limited knowledge of Spanish. Expressive language skills were obtained via a narrative story telling task using two books; done once in English for the monolinguals (book choice was randomized) and done twice for bilinguals, once in English and once Spanish (book choice and order were randomized). After a short break, children then completed three executive functioning tasks [e.g., the Dimensional Change Card Sort (DCCS), the Simon task, a go/no-go task (GNG)], followed by a second break, and ended with two more executive functioning tasks [e.g., Tower of London (TOL) and the Head-Toes-Knees-Shoulders task (HTKS)].

Measures

Receptive Language Skills

The Peabody Picture Vocabulary Test (PPVT-III; Dunn & Dunn, 1997) was used to directly assess children's receptive language skills in English, which can be administered to individuals ranging in age from 2½ years through adulthood. Children are presented with a "picture plate" that contains four pictures per plate, and are then asked to point to a target word. Raw scores are converted into standard scores by using norms tables based on a standardization sample of 2,275 individuals. The split-half reliability coefficients for the PPVT-III, based on the standardization sample, ranges from .86 to .97 (Kaufman & Lichtenberger, 2006), and the PPVT-III has been found to correlate highly with other measures of intelligence, indicating good construct validity. The Test de Vocabulario en Imágenes Peabody (TVIP; Dunn et al., 1986) is the Spanish version of the PPVT-Revised (Dunn & Dunn, 1981), and is administered in the same manner as the PPVT. The TVIP was normed on samples of Spanish-speaking monolingual children in Mexico and Peru, and the median split-half reliability coefficient was .93.

Executive Functioning

Parents completed the Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000), which is an 86-item questionnaire (based on a three-point scale; never, sometimes, often) that measures child behavioral regulation and metacognitive skills. The BRIEF produces a Global Executive Composite, in which higher scores indicate greater impairment in executive functioning, and this measure has demonstrated high internal consistency (.82-.98; Gioia et al., 2000). Children were administered a total of five direct assessments of executive functioning, aimed to measure cognitive flexibility, interference control, inhibition of a prepotent response, planning, and inhibitory control. To assess cognitive flexibility, the Dimensional Change Card Sort (DCCS; Frye, Zelazo, & Palfai, 1995) was used, which requires children to first sort cards by one sorting rule (Block 1, shape), and then sort by a new rule (Block 2, color), when instructed by the experimenter. A third sorting rule (Block 3) was used in this study, such that when a star was on the card, it should be sorted by color, but when there was no star, the card should be sorted by shape. The more challenging third block has been suggested for use in children up to age 7 (Zelazo, 2006), which is appropriate for the age range of the current sample. The scores for this measure reflect the number of errors made during a given block (i.e., more errors indicate poorer performance), and due to a ceiling effect found in Blocks 1 and 2, the errors made in Block 3 were used in further analyses. The DCCS has been used with bilingual samples (Bialystok, 1999; Bialystok & Martin, 2004), and has convergent validity as indicated by correlations with academic skills and teacher-report ratings of child behavior (*r*s between .42-.63) and predictive validity on later academic skills (.43-.64; Lipsey, Wilson, & Farran, 2010.

Interference control was measured using the Simon task, which has been used with bilingual samples (Bialystok, Craik, Klein, & Viswanathan 2004; Martin-Rhee & Bialystok, 2008). During this task, individuals must push a button on one side of a keyboard (e.g., a key on the left side) when one stimulus is presented on a screen, and another button (e.g., a key on the right side) must be pressed when a second stimulus appears on the screen. There are two possible presentations: congruent trials (stimulus appears on the same side of the screen as the appropriate button) and incongruent trials (stimulus appears on the opposite side of the screen from where the designated button is on the keyboard). Scores from this task included reaction time and percent accuracy for each trial type (congruent and incongruent), such that lower scores indicate faster and/or more accurate performance. The Simon Effect was calculated by subtracting the reaction time of the congruent trials from the incongruent trials, such that smaller numbers reflect better interference control. The Simon Effect has been found to be robust and reliable (Simon, 1990), thus, the Simon Effect score was used in further analyses.

The go/no-go (GNG) task was used to measure inhibition of a prepotent response, which has been used with bilinguals (Emmorey, Luk, Pyers, & Bialystok, 2008). The task requires children to push the spacebar as fast as possible when squares with straight or diagonal lines appeared on the screen ("go" stimulus), but not when a square with an X appeared ("no go" stimulus). Scores from this task include errors of commission (making a response to the "no-go" stimulus) and errors of omission (not making a response on the "go" stimulus). The total number of errors of commission were used in further analyses because this component is interpreted as a measure of behavioral inhibition (Berlin & Bohlin, 2002). The GNG task has good test-retest reliability (inter-class correlations .5-.9; Kuntsi, Andreou, Ma, Borger, & van der Meere, 2005).

For planning, the Tower of London (TOL; Shallice, 1982) was used, which involves a set of different colored balls on three wooden pegs of different lengths, and the child is asked to match his/her set to the experimenter's set in the fewest moves possible.

Performance on the TOL was measured in two ways: latency time (the time from the end of the instructions until the child's first move), and the total move-value (the total number of moves made during the correctly-solved trials). Due to a methodological error in the administration of the TOL task, balanced bilinguals and monolinguals had unintentionally longer times spent in the instruction portion of the task, so the latency time was considered invalid. Instead, only the total move-value was used, such that higher scores indicated better performance on the task. As reported in the original study, there was high inter-rater reliability for coding the number of moves during the TOL task (r = .99; Hutchison, 2012).

Finally, the Head-Toes-Knees-Shoulders (HTKS) task (Matthew, Ponitz, & Morrison, 2009; Ponitz et al., 2008) was administered to measure inhibitory control, such that the experimenter asks the child to touch his/her toes, but the child is required to do the opposite (touch his/her head). The HTKS was scored as the following: two points for a completely correct response, one point for a self-corrected response, and zero points for a completely incorrect response. The total score on the HTKS could range from 0 to 40 points, with higher scores reflected better performance, and HTKS has demonstrated construct validity with parent- and teacher-report ratings of child behavior (Ponitz et al., 2008).

Narrative Storytelling

Following the elicitation procedures outlined by the Narrative Assessment Protocol (NAP; Justice et al., 2010), children were instructed to go through the pictures in a book, and then create a story based on the pictures of the book while making the story as long and interesting as possible. All narratives were transcribed word-for-word into Word documents by one research assistant (RA) and were then verified by a second RA. Seven bilingual RAs transcribed and verified the Spanish narratives, and six monolingual RAs transcribed and verified English transcripts according to the Codes for the Human Analysis of Transcripts (CHAT: MacWhinney, 2000), which is a standardized set of symbols and codes used to translate verbal cues into a computer text format. (For complete details on the transcription procedures, please refer to Hutchison, 2012). There is support that individuals can be reliable in accurately transcribing oral narratives in English Language Learners (Heilmann et al., 2008).

Two books, *Frog, Where Are You?* (Mayer, 1969) or *Frog Goes to Dinner* (Mayer, 1974), were used during this story retell task. The *Frog* books are long, multiple-episode story books with detailed scenes and facial expressions. According to Lofranco et al. (2006), narratives produced based on the *Frog* book were more complex compared to two single-episode books. Considering the complexity of the *Frog* books, the *Frog* books could be considered developmentally appropriate for young children because they provide multiple opportunities to identify and describe episodes, which may be sensitive to detect age-related differences. The choice (for all children) and/or order (for bilinguals) of the *Frog* books was randomized across participants. Hutchison (2012) found no main effect for the type of book used or the order of language presentation (first English then Spanish or vice versa) on word production.

Coding of Narratives

Three monolingual graduate students coded the English transcripts, and four bilingual undergraduate and graduate students coded the Spanish narratives. For both the English and Spanish coding, two students independently coded several transcripts containing 15-20% of the total number of narratives in order to achieve inter-rater reliability. For dichotomous codes, inter-rater reliability was examined using Kappa and percentage agreement, and Spearman-rank correlations was used for ordinal codes.

Microstructure of Narratives

Language sample indices. Language sample indices capture the use of expressive language skills during a narrative speech sample, such as the number of total words (NTW), number of different words (NDW), and mean length of utterance (MLU). For total number of words, any and all intelligible words the child spoke were part of this code category, including grammatical errors, repeated words, and restarts. Words or 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) and any filler words like "um" or "uh" were excluded from analyses. Inter-rater reliability was high for both total word count in both English (r = .99) and Spanish (r = 1.00). For NDW, unique words that the child uttered across the entire transcript were identified (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). Inter-rater reliability was also high for unique/different word count in English (r = .97) and Spanish (r = .98). MLU was calculated by taking total the number of words and dividing that by the total number of utterances. Based on prior research using language sample indices, there is a risk for multicollinearity if all three variables (i.e., NTW, NDW, and MLU) are included together due to the high correlation among these language indices (Justice et al., 2006), so one index had to be selected for further analyses. The number of unique words is suggested to be more reflective of a child's language growth, compared to the total number of words in narrative complexity, thus, the number of unique words was used (Bedore et al., 2010; Uccelli & Páez, 2007).

Narrative Assessment Protocol (NAP). Justice and colleagues (2010) developed a coding system called the Narrative Assessment Protocol, which analyzes the microstructure of children's narratives generated from the telling of a *Frog* wordless picture book. In the original version of the NAP, 18 different language forms are organized into five indicators: sentence structure (e.g., compound sentence), phrase structure (e.g., elaborated noun phrase), modifiers (e.g., adverbs), nouns (e.g., plural noun), and verbs (e.g., irregular past tense). Each linguistic form is then coded for frequency based on a 0 to 3 rating scale (i.e., 0 = did not occur, 3 = occurred 3+ times), with a possible score range from 0-54. The 18-item version is now referred to as the Long Form (in its original development), and a 12-item Short Form version was created after six items were considered to not load strongly enough onto the one-factor solution based on an exploratory factor analysis (Justice et al., 2010), with a possible score range from 0-36. Since its publication, a growing number of studies have utilized the short form (Gosse, McGinty, Mashburn, Hoffman, & Pianta, 2014; Justice, McGinty, Zucker, Cabell, & Piasta, 2013; Terry et al., 2013), so the NAP Short Form score was used in this study. The NAP offers online reliability training and access to coding sheets. The lead author used the online training materials to become reliable, which includes gaining familiarity with the coding procedures on two videos, practicing using the coding sheets on five videos, and then independently coding three videos for reliability to be within one point agreement with the master codes on 15 out of 18 items. The lead author was found to be reliable on the first pass of three videos, such that the lead author matched the master codes within one point on 17 out of 18 items on the first video, 18 out of 18 on the second and third videos.

Attempts have been made to create a Spanish adaptation of the NAP is in development (Gorman, Bingham, Fiestas, & Terry, 2015; Solari et al., 2013). The Spanish adaptation by Solari and collegues (2013) included all English codes from the original NAP, but also included three additional Spanish codes to account for the different linguistic forms not found in English. For example, the use of definite articles (e.g., *el*, *la*, *los*), clitic pronous (similar to English pronouns, but must agree in number and gender), and pluralized pronouns (pronouns that are pluralized and act as nouns). This version has yet to be published, but shows promise for use in Spanish-speaking children. A new Spanish adaptation of the NAP is under development and is available for use with permission from the lead author (Gorman et al., 2015). This version includes nine additional items to capture Spanish-specific linguistic forms, and results have indicated that the overall total score reflects one underlying construct, however, several items did not significantly load onto the factor (Gorman et al., 2015). This version demonstrates test-retest reliability between fall and spring scores in preschoolers, and concurrent validity with scores on a standardized expressive vocabulary test. By applying this version of the coding system to a new sample of varying degrees of bilingual children, the current study provides useful information on the validity of the measure. While the NAP was originally developed for use in preschoolers, it may be useful for coding narratives in Dual Language Learners who may exhibit language delays in one or both languages (Pearson et al., 1993).

The NAP was developed to provide researchers with developmentally sensitive measures of children's narratives. In Justice and colleagues' (2010) original study on the NAP, the authors set out to develop a standardized measure of narratives that can be assessed without the need to transcribe the language sample. The materials include the cost-effective and accessible *Frog Where are You?* book used to elicit the narrative, coding sheets that are available on the researchers' website, as well as the training materials needed for future studies. The reliability between on-site and online training was found to be similar, suggesting that outside researchers can be reliably trained using their online resources. For the 3- to 5-year-old children from the original sample, the NAP was reported as demonstrating strong construct and concurrent validity, and predictive validity with standardized language measures (Justice et al., 2010). Terry et al. (2013) coded children's narratives using four measures of narrative microstructure and macrostructure across the pre-K school year, and the authors praised the NAP for being capable of detecting change over the pre-K year, practical for clinical use (due to the live coding without transcription), and highly correlated with coding schemes that focus on narrative macro-structure.

When a Spanish-adaptation of the NAP was assessed with low-SES, English Language Learners, the authors were unable to compare English and Spanish NAP scores because of significant floor effects of the English NAP scores (Solari et al., 2013), and the Spanish version did not yield the same factor structure as the original, English publication. In Gorman et al.'s adaptation (2015), the solution did yield a similar factor structure and the assessments of validity showed more promise, so this adaptation was used in this study. In the current sample, a factor analysis was run to examine the factor structure of Gorman et al.'s (2015) adaptation of the NAP-Spanish. Based on the eigenvalues and visual inspection of the scree plot, there yielded one factor (eigenvalue = 11.38), and all items, except for the subjunctive verb code, loaded onto Factor 1. When all 27 items were included, Cronbach's alpha was high, $\alpha = .94$.

Macrostructure of Narratives

Story Grammar (SG). Story grammar is one of the earliest coding systems used with narratives, and it focuses on the macrostructure of the story. In this coding system, there are seven story elements: setting (introduction of characters and the context), initiating event (the occurrence of an event that causes the main character to respond), internal response (the characters' thoughts and feelings event), plan (intended actions in order to reach goal), attempt (actions to the pursuit of a goal), consequence (attainment or non-attainment of the goal), and ending (resolution of the problem) (Bamberg, 1987; Stein & Glenn, 1979; Thorndyke, 1977). First, the total frequency of each of the seven story grammar element (i.e., setting, initiating event, internal response, plan, attempt, consequence, ending) was calculated and then aggregated to yield a total number of story grammar elements identified in a story. Second, a dichotomous 0/1 code was assigned to each story grammar element to identify whether or not a child ever used an element in a story. For example, if a child produced at least one setting statement but no planning statements, he/she would get a 1 for "setting" and a 0 for "planning." These 0/1 codes were then aggregated to reflect how many of the seven elements were identified in a story. Using the 0/1 scoring, there is a narrow range of possible scores (min = 1, max = 7), the total number of story grammar elements was used for further analyses because of the greater amount of variability in this variable found in this study (min = 1, max = 36).

Story grammar coding has been used with young school-age children, including monolinguals (Schachter & Craig, 2013) and bilinguals (de Quirós et al., 2012; Fiestas & Peña, 2004, Muñoz et al., 2003), as well as with adults (Coelho et al., 1990). With the inclusion of sophisticated story elements in addition to the minimum requirements to be classified as a complete episode, this coding system is sensitive to developmental changes in children's narratives (Muñoz et al., 2003).

High-Point Analysis (HPA). High-Point Analysis is based on a single, 0-7 point scale based on the overall structure and quality of a narrative. As the name implies, the high-point (or climax) is the most advanced element of a well-structured and organized story (McCabe, 1997). A "classic" narrative (worth 7 points) provides details on the characters and setting of the events, presents a series of events that build up to a high point and evaluates the situation, and ends with a resolution (McCabe, 1997; McCabe et al., 2008). If there is no resolution, the narrative would be scored a 6, and if the events are mostly presented chronologically without a high-point, then it would be scored a 5, and so on. A narrative would score a 0 if the events are described in the present tense, with no reference to past or future events, and contains only descriptions of the scenario. HPA was originally designed to evaluate personal narratives (McCabe, 1997), but has also been used for narrative storytelling (McCabe et al., 2008; Terry et al., 2013) in addition to personal narratives (Bailey et al., 2008). High-Point Analysis (HPA) is a coding system than has been used for many years and across various ages (Bailey et al., 2008; McCabe et al., 2008; Terry et al., 2013). HPA is described as better able to capture coherence of a story compared to story grammar, which addresses how well individuals plan their story (McCabe, 1997).

Inter-Rater Reliability

English

For English coding, the number of total words and unique words were reliably coded by two monolinguals as reported in the original study (Hutchison, 2012). For the HPA and story grammar, two graduate students independently coded a minimum of 15% of the transcripts In order to achieve inter-rater reliability for each coding system. For HPA reliability, 12 transcripts were coded for reliability, and the Spearman correlation was $r_s = .74$, $p < .01$, and 10 out 12 (83.3%) of the codes were within 1 point of each other. The reliability statistics for story grammar in English were adequate for setting ($\kappa = .77$, 97.2% agreement), initiating event ($\kappa = .65$, 93.7%), internal response ($\kappa = .93$, 94.1%), attempt ($\kappa = .66$, 92.1%), and ending ($\kappa = .65$, 96.4%) statements. For planning, the Kappa value was low ($\kappa = .50$, 98.8%) due to a low frequency of planning statements, and the Kappa value for consequences was also low ($\kappa = .51$, 81.8%); however, additional training took place before proceeding with the remainder of the coding. It is important to consider that while the Kappa values were low in some areas—likely due to a low base-rate of frequency—Kappa values are not typically reported in studies that use story grammar. More specifically, percent agreement is used as an index of reliability, for example, 93% overall agreement was reported in a study of monolingual English speakers (Schachter & Craig, 2013). Therefore, while the Kappa values were low, the percent agreements were high enough to be considered reliable.

Spanish

For Spanish coding, the number of total words and unique words were reliably coded by two bilingual, English/Spanish speakers as reported in the original study (Hutchison, 2012). Two new bilingual undergraduate students were trained and independently coded 10 transcripts for HPA and story grammar reliability. For HPA, the Spearman correlation was $r_s = .84$, $p < .01$, and 9 of out 10 (90.0%) of the transcript codes were within 1 point of each other. The reliability statistics for story grammar were adequate for setting ($\kappa = .80$, 98.9% agreement), initiating event ($\kappa = .70$, 93.7%), internal response ($\kappa = .66$, 98.9%), planning ($\kappa = 1.00$, 100.0%), and ending ($\kappa = .65$, 96.4%) statements. The Kappa values for attempts ($\kappa = .55$, 92.6%) and consequences ($\kappa = .50$, 97.9%) were low; however, additional training took place before proceeding with the remainder of the coding. As mentioned above, studies that include story grammar report percent agreement as an index of reliability, with values ranging from 90-91% agreement in bilingual samples (de Quirós et al., 2012; Fiestas & Peña, 2004; Muñoz et al., 2003). Therefore, while the Kappa values were low, the percent agreements were high enough to be considered reliable. For the Spanish NAP, two bilingual students (one undergraduate and one graduate student) independently coded three transcripts for reliability, in an attempt to become reliable using the same standards as the English version. More specifically, the two bilingual coders gained familiarity with the training procedures and practiced coding three transcripts in English, and were then given three Spanish transcripts to code. The same criteria for Spanish was used as the English system, such that coders had to be within-one point agreement on 83.3% of the codes. The within-one point percent agreements for the first three transcripts were 75.0%, 95.0%, and 90%. Following the reliability procedures for English, an additional three transcripts were coded and all three met the reliability threshold (85.0%, 90.0%, and 90.0% within one-point percent agreement).

RESULTS

Descriptive Statistics

Exploratory data analyses (EDAs) were run for each component of each measure of narrative complexity in order to determine the amount of variability and the level of normality of each component to aid in the selection of the components for use in further analyses. While some individual NAP items and story grammar elements were slightly skewed, overall, the narrative measures were found to be mostly normally distributed with few identifiable outliers. For example, the total number of words produced in English narratives did have one large outlier, but this did not affect further analyses because this variable was not used. There were also no bivariate outliers. EDAs were also run separately by language group, and for the English narrative complexity scores, there were very similar in terms of means and variability measures. For example, if there were one large value for the monolinguals, there was one individual with a similar value in both the balanced bilingual group and DLL group. Thus, the descriptive statistics for English narrative stories include the entire sample in Table 1, and this table also contains the descriptives for the Spanish stories for the subsample of the DLL and balanced bilingual children. Scatterplots were also run to identify any bivariate outliers with all narrative measures plotted against all background variables and executive functioning variables, and there were no identifiable bivariate outliers found, neither in English nor in Spanish.

Table 1. Descriptive Statistics on Narrative Complexity Measures (English & Spanish)

	English (*n* = 76)				Spanish (*n* = 35)			
	Mean	(*SD*)	Min.	Max.	Mean	(*SD*)	Min.	Max.
Language Sample Indices								
Number of Total Words	336.45	(172.74)	131	1,145	212.97	(143.72)	9	713
Number of Unique Words	102.82	(36.57)	57	268	58.69	(39.26)	1	193
Number of Utterances	33.07	(16.01)	19	123	26.81	(10.40)	5	53
Mean Length of Utterance (MLU)[a]	10.25	(2.67)	5.33	17.69	7.52	(4.14)	1.13	17.39
Narrative Assessment Protocol								
English Short Form (12 items)	23.86	(4.70)	14	35				
Spanish Total (27 items)					19.81	(16.88)	0	53
Story Grammar								
Total Number of Story Elements	18.16	(7.03)	3	36	7.06	(5.07)	0	22
1/0 Usage of 7 Story Elements	5.32	(1.11)	2	7	4.22	(2.14)	0	7
High Point Analysis	5.68	(1.16)	3	7	3.89	(2.39)	0	7

[a]MLU is calculated by dividing Number of Total Words by Number of Utterances

English

For the language sample indices in English, the total number of words and number of unique words were reported in the original study (Hutchison, 2012), and the total number of utterances and mean length of utterance (MLU) were tallied and calculated for the current study. On average, across all language groups, an English story consisted of roughly 336 total words (102 of which were unique words) across 33 utterances, thus yielding an average MLU of 10 words per utterance. The Short Form of the NAP yielded a mean score 23.86 (*SD* = 4.70) revealed a normal distribution and wide range of scores (relative to the maximum possible score). For story grammar, on average, a child would use roughly 18 story elements per story, but the range varied from only 3 elements to 36 elements. For the 0/1 codes of how many of the seven elements were identified in a story, an average of five elements were produced per story. For the High

Point Analysis (HPA), the possible range of scores was from 0 to 7, but for this sample, there was a restricted range of 3 to 7. In other words, none of the children in this sample produced an English story that could be coded as anything lower than 3 on the ordinal HPA scale. Thus, the distribution of HPA codes was slightly negatively skewed, with most children's stories coded as either a score of 5 or 6, and fewer stories classified as 3-4 or 7.

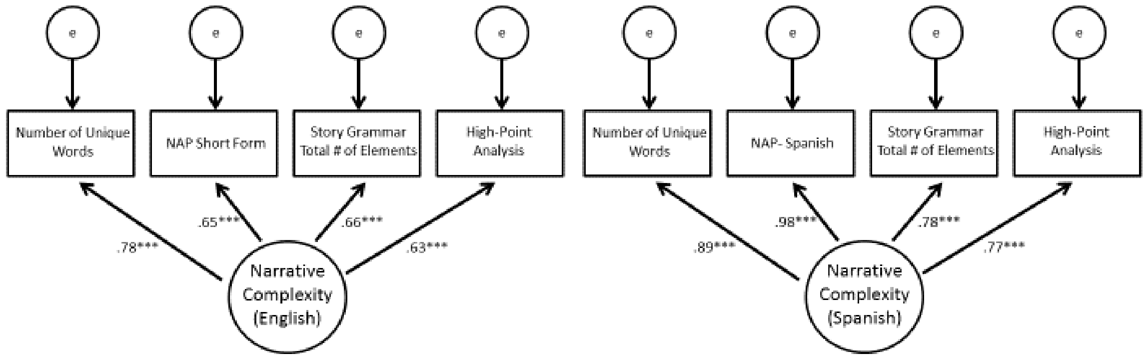
Spanish

On average, across the two bilinguals groups, a Spanish story consisted of roughly 213 total words (58 of which were unique words) across 26 utterances, thus yielding an average MLU of seven words per utterance. The total number of points on the NAP-Spanish yielded a mean score 19.81 (*SD* = 16.88) and revealed a somewhat normal distribution and wide range of scores. For story grammar, on average, a child would use roughly seven story elements per story, but the range varied from no elements to 22 elements. For the 0/1 codes of how many of the seven elements were identified in a story, an average of four elements were produced per Spanish story. For the High Point Analysis (HPA), there was a full representation of scores in the possible range of scores (i.e., 0 to 7) and the distribution of scores was relatively normal. Compared to the English stories, the Spanish stories had two-thirds of the total words produced, and almost half of the number of unique words during the storytelling task. The ranges of scores on each of the language sample indices revealed that all Spanish stories were shorter and contained fewer total and unique words compared to the English stories. For the NAP, story grammar, and HPA, there was more variability (i.e., larger *SD*s) in the stories produced in Spanish compared those produced in English.

Exploratory Factor Analyses

Narrative Complexity

Two exploratory factor analyses (EFAs) were conducted—once in English, once in Spanish—to determine if an underlying construct of narrative complexity could be found among the four different measures of narrative complexity (i.e., language sample indices [number of unique words, number of total words, MLU], NAP, story grammar, high-point analysis). For English, using Mplus, an initial model was run with the number of unique words, the NAP short form score, the total number of story grammar elements, and the HPA score, freeing the factor variance of the latent construct. Using the recommendations of indicators of good fit from Hu and Bentler (1995), overall, the initial model had excellent fit (see Figure 1a). More specifically, the chi-square test of model fit was χ^2 (2) = 1.90, p = .387, indicating that my observed correlation matrix is not significantly different from the predicted correlation matrix. Additional fit indices also revealed a good-fitting model, RMSEA = .00 and CFI = 1.00. As presented in Figure 1a, each of the four parameter estimates loaded significantly onto the latent construct. A second model was run to include the total number of words and mean length of utterance (MLU), but the results indicated a poorly-fit model largely due to the large correlations among the language sample indices, so the model presented in Figure 1a was used in subsequent analyses (i.e., the number of unique words, the NAP short form score, the total number of story grammar elements, and the HPA score).



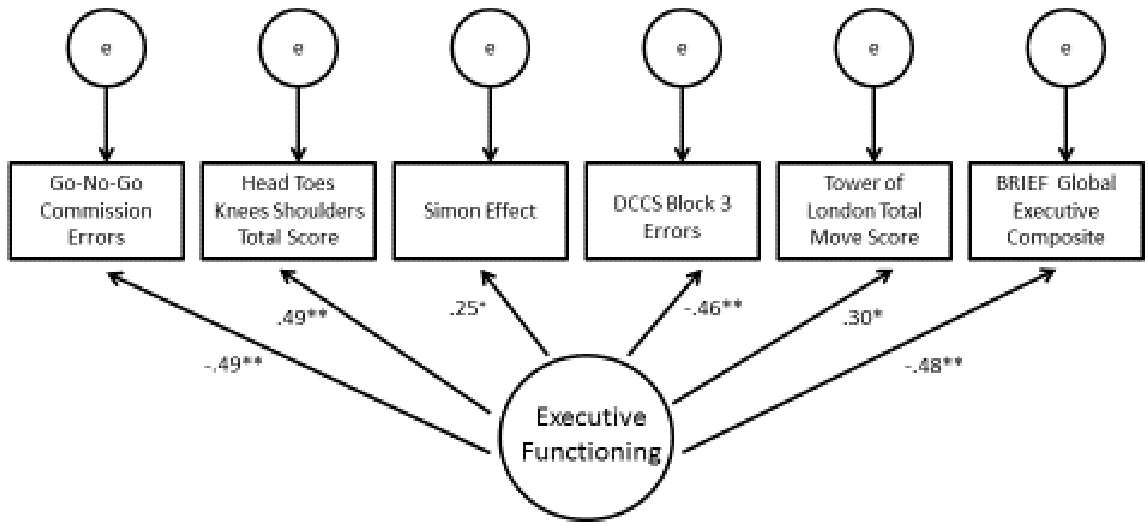
Notes. **p* < .05, ***p* <.01, ****p* < .001

Figure 1. Standardized Results of Exploratory Factor Analysis of Narrative Complexity (1a. English and 1b. Spanish)

Similarly, for Spanish, using Mplus, an initial model was run with the number of unique words, the NAP- Spanish total score, the total number of story grammar elements, and the HPA score, freeing the factor variance of the latent construct. Overall, the initial model had good fit (see Figure 1b). More specifically, the chi-square test of model fit was χ^2 (2) = 5.31, p = .070, and additional fit indices also revealed poor- to good-fitting model, RMSEA = .21 and CFI = .97. As presented in Figure 1b, each of the four parameter estimates loaded significantly onto the latent construct. Compared to the English EFA, the standardized parameter estimates loaded more strongly onto the Spanish narrative latent factor (i.e., parameter estimates ranging from .77-.98) than the English narrative latent factor (i.e., parameter estimates ranging from .63-.78).

Executive Functioning

An exploratory factor analysis was conducted to determine if an underlying construct of executive functioning could be found among the areas measured of executive functioning (i.e., cognitive flexibility, interference control, inhibition of a prepotent response, planning, inhibitory control). Using Mplus, an initial model was run with the number of commission errors from the Go-No-Go task, the total score from the Head Toes Knees Shoulders task, the Simon Effect, the number of errors in the third block of the Dimensional Change Card Sort task, the total move value from the Tower of London task, and the BRIEF Global Executive Composite. The path of the Head Toes Knees Shoulders score was fixed to 1 in order for the latent construct to reflect larger, more positive scores associated with better executive functioning performance. Overall, the initial model had excellent fit. More specifically, the chi-square test of model fit was χ^2 (9) = 8.41, p = .493, and additional fit indices also revealed a good-fitting model, RMSEA = .00 and CFI = 1.00. As presented in Figure 2, five of the six parameter estimates loaded significantly onto the latent construct, Simon Effect yielding a marginally significant factor loading.



Notes. +p <.10, *p < .05, **p <.01

Figure 2. Standardized Results of Exploratory Factor Analysis of Executive Functioning

An additional model was run to control for age, and the resulting model indicated good fit, χ^2 (14) = 18.60, p = .181, and additional fit indices also revealed an adequate-fitting model, RMSEA = .07 and CFI = .88. The parameter estimates were all similar in magnitude, direction, and level of significance compared to the initial model, and age significantly predicted overall executive functioning, β = .58, p < .001. In other words, as the age of the child increases, the overall performance on executive functioning also increases. Pearson correlations were also run to determine how each background variable related to each outcome variable (see Table 2). Age was negatively correlated with the number of errors on both the Go-No-Go task and the DCCS task (*r*s -.34 and .40 respectively), and English receptive language was also negatively correlated with the number of DCCS errors (r = -.26). Based on these results, age was used as a covariate in further analyses that included executive functioning.

Table 2. Pearson Correlations between Background Variables and All Measures

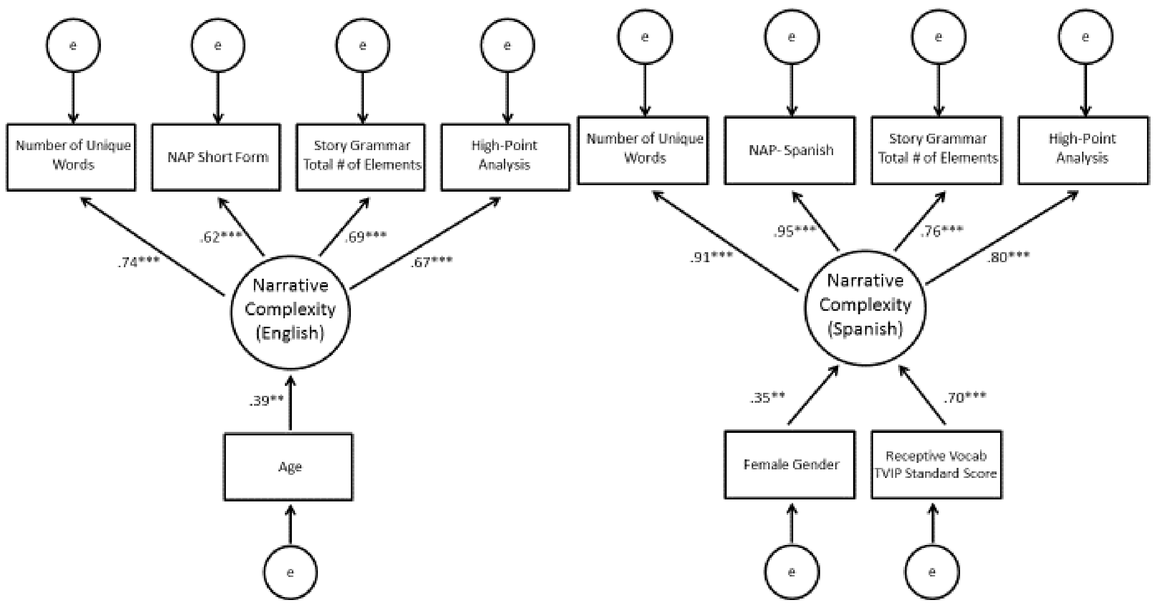
	English (n = 76)			Spanish (n = 35)		
	Age (Months)	Female Gender	Receptive (PPVT)	Age (Months)	Female Gender	Receptive (TVIP)
Narrative Measures						
# of Unique Words	.19	$.33^{**}$	.17	.24	$.37^{*}$	$.65^{***}$
NAP	.13	.13	.05	.15	$.31^{+}$	$.67^{***}$
SG (# of Elements)	$.34^{**}$	.04	.04	$.51^{**}$	$.40^{*}$	$.34^{*}$
High Point Analysis	$.39^{***}$	.08	$.21^{+}$	$.29^{+}$	.17	$.71^{***}$
EF Measures						
GNG Commission Errors	$-.34^{**}$	$-.20^{+}$	-.04	$-.41^{*}$	$-.32^{+}$	-.15
HTKS Total Score	.17	.05	-.04	.19	-.10	-.13
Simon Effect	.10	.07	-.05	.02	-.14	-.15
DCCS Block 3 Errors	$-.40^{***}$	.09	$-.26^{*}$	$-.57^{***}$	.12	-.22
TOL Total Move Score	$.21^{+}$	.01	-.10	.25	.00	-.19
BRIEF Global Ex. Composite	.01	-.08	-.10	-.08	-.16	-.05

$^{+}$ $p < .10$, *$p < .05$, **$p < .01$

Question 1. What are the predictors of narrative complexity?

English

Based on the results of the EFA for narrative complexity in English, a structural equation model was tested that included three predictors (i.e., age, gender, receptive language skills) in the model; however, the results indicated poor fit, χ^2 (11) = 25.08, $p < .01$, RMSEA = .13, and CFI = .85, and receptive vocabulary (i.e., PPVT) and gender did not significant predict narrative complexity. A second model was run to only include age as a predictor, and the resulting model indicated good fit, χ^2 (5) = 9.25, p = .100, and additional fit indices also revealed an adequate (RMSEA = .10) to good-fitting model (CFI = .95). The parameter estimates (see Figure 3a) for the predictors revealed that age significantly predicted overall narrative complexity. In other words, as the age of the child increases, the story the child told in English was more complex. Pearson correlations were also run to determine how each background variable related to each narrative measure (see Table 2). The results indicated that for English, story grammar and HPA were both positively correlated with age (*r*s .34 and .39 respectively), the number of unique English words was positively correlated with female gender (r = .33), and receptive vocabulary in English (i.e., PPVT) was marginally correlated with HPA (r = .21).



Notes. +p <.10, *p < .05, **p <.01, ***p < .001

Figure 3. Standardized Results of Predictors of Narrative Complexity (3a. English and 3b. Spanish)

Spanish

Based on the results of the EFA for narrative complexity in Spanish, a structural equation model was tested that included all three predictors (i.e., age, gender, receptive language skills) in the model. The results, however revealed a very poor fit, χ^2 (11) = 47.43, $p < .001$, RMSEA = .30, and CFI = .79, and child age was the only predictor to not load significantly onto the latent construct. A second model was run excluding age as a predictor, and the resulting model indicated poor fit, χ^2 (8) = 19.32, $p < .05$, RMSEA = .19, although the CFI = .92 was adequate. The parameter estimates (see Figure 3b) for the predictors revealed that both gender and receptive language (i.e., TVIP) significantly predicted overall narrative complexity. In other words, as the receptive vocabulary in Spanish of the child increases, the story the child told in Spanish was more complex, and girls generally produced more complex narratives compared to boys. Pearson correlations between each background variable and each narrative measure (see Table 2) indicated that for Spanish, the total of story grammar elements was positively correlated with age (r = .51), the number of unique Spanish words and story grammar elements were positively correlated with female gender (rs .37 and .40 respectively), and receptive vocabulary in Spanish (i.e., TVIP) was correlated with all Spanish narrative measures (rs ranging from .34-.71).

Question 2. In bilingual children, are there language differences (between English and Spanish) in narrative complexity?

A series of paired sample t-tests was run by comparing the narrative complexity scores (the DVs) across English and Spanish (the IV: language of assessment) for all bilingual children (n = 35). Table 3 provides the means and standard deviations for each narrative assessment in each language. The bilingual children in this sample produced significantly more unique words in English than in Spanish, $t(34) = 8.18$, $p < .001$, significantly more story grammar elements in English than in Spanish, $t(34) = 9.47$, $p < .001$, and significantly higher scores on the High Point Analysis in English than in Spanish, $t(34) = 4.38$, $p < .001$. There was, however, no difference between the NAP-Short Form in English and the NAP-Spanish version, $t(34) = 1.29$, $p = .206$.

Table 3. Language Differences between English and Spanish for Bilingual Groups (Controlling for Receptive Vocabulary TVIP)

	All Bilinguals (n = 35)			Balanced (n = 15)			DLLs (n = 20)		
	Mean	(*SD*)	*r*	Mean	(*SD*)	*r*	Mean	(*SD*)	*r*
Number of Unique Words[abc]			.56**			.88***			.26
English	109.09	(36.85)		112.93	(45.90)		106.20	(29.29)	
Spanish	58.97	(39.79)		86.87	(37.04)		38.05	(27.34)	
NAP[bc]			.06			.54*			-.03
English	24.03	(4.84)		23.40	(3.94)		24.50	(5.48)	
Spanish	20.26	(16.91)		32.60	(13.57)		11.00	(12.86)	
Story Grammar (Total # of Elements)[abc]			.30^{+}			.26			.47*
English	19.66	(7.66)		19.60	(7.23)		19.70	(8.15)	
Spanish	7.17	(5.09)		10.07	(5.37)		5.00	(3.69)	
High Point Analysis[ac]			.25			.01			.30
English	5.69	(0.99)		5.80	(0.86)		5.60	(1.10)	
Spanish	3.94	(2.40)		5.40	(1.18)		2.85	(2.52)	

$^{+}$ $p < .10$, *$p < .05$, **$p < .01$, *** $p < .001$

[a] *significant difference between English and Spanish for all bilinguals*

[b] *significant difference between English and Spanish for balanced bilinguals*

[c] *significant difference between English and Spanish for DLLs*

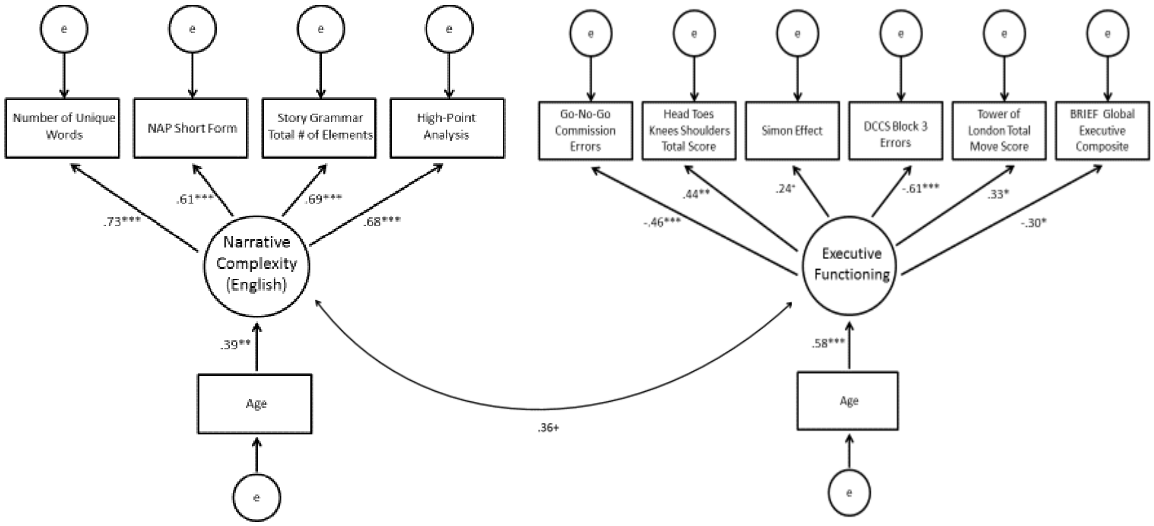
Additional analyses were run to explore if the language differences between English and Spanish were the same for the DLL group and the balanced bilingual group (see Table 3). When re-running the analyses separately by group, results indicated that the DLLs in this sample had higher scores in the English narratives compared to the Spanish narratives on all narrative assessments, and the balanced bilingual group had similar results, although the HPA scores were similar in English and Spanish.

Correlations between the English and Spanish narrative were also run across all bilingual children, and then separately for the DLL and balanced bilingual groups. As presented in Table 3, there was a significant positive correlation (r = .56) between the number of unique words in English and Spanish across all bilingual children, and this correlation was especially strong within the balanced bilingual group (r = .88). At first, it appeared that there was no correlation between the languages on the NAP, but upon further inspection, there was a positive correlation for the balanced bilingual group (r = .54) while there was almost no relation for the DLLs (r = -.03). It is possible that the skewed nature of the NAP-Spanish for the DLLs may have contributed to this zero correlation, such that the standard deviation was similar in magnitude as the mean. Overall, there was a marginal correlation between the number of story grammar elements included in English and Spanish stories, and this was significant for the DLL group. These results indicated that there was a strong positive correlation between English and Spanish on some narrative measures, while there was close to a zero correlation in other narrative measures (i.e., NAP for DLLs, HPA for balanced bilinguals). There were no significant correlations between the two languages on the HPA, which is similar to a previous study that examined English and Spanish narratives in a sample of Spanish-dominant DLLs (Bailey et al., 2008).

Question 3. How does narrative complexity relate to executive functioning across all children?

English

Based on the results of the EFAs for both narrative complexity in English and executive functioning, a structural equation model was run in Mplus to determine if there were a relation between narrative complexity and executive functioning. The results revealed a model that had good fit, χ^2 (34) = 39.54, $p = .236$, and additional fit indices also revealed a good- to adequate-fitting model, RMSEA = .05 and CFI = .95. There was a strong positive correlation ($r = .49$, $p < .01$) between the latent constructs of narrative complexity and executive functioning. Thus, children who produced English stories with greater complexity also had better performance on executive functioning assessments. A second model was run to include age as a predictor for both narrative complexity and executive functioning, as found from the results of Question 1 and the EFAs. The resulting model had slightly poorer fit, χ^2 (42) = 53.32, $p = .113$, RMSEA = .06 and CFI = .91, but the parameter estimates were all similar in magnitude and level of significance (see Figure 4). Controlling for age, there remained a positive correlation between the latent constructs of narrative complexity and executive functioning, but this was only found to be marginally significant given the sample size and loss of degrees of freedom ($r = .36$, $p = .06$).



Notes. +*p* <.10, **p* < .05, ***p* <.01, ****p* < .001

Figure 4. Standardized Results of Relation between Narrative Complexity (in English) and Executive Functioning

In addition to the structural equation model, Pearson correlations were run among each narrative complexity score and each executive functioning task score, controlling for age (see Table 4). There were no significant correlations between the number of unique words or the number story grammar elements and any EF task. There was only one significant positive correlation between the NAP short form and the HTKS task total score. In other words, children who had a higher score on the NAP short form earned more points (i.e., better performance) on the HTKS task. For HPA, a marginally negative correlation with the DCCS errors was found, as well as a significant positive correlation with the TOL total move score, such that children who scored higher on the HPA solved the TOL task in the fewest moves possible (i.e., better performance), and also made marginally fewer errors on the most challenging block of the DCCS task. Pearson correlations were also run without controlling for age, and the results remained the same with two exceptions: the positive correlations between DCCS errors and both HPA and story grammar became significant. These results indicated that there were a handful of meaningful correlations between each individual narrative measure and the Head Toes Knees Shoulders task and the Tower of London task, while there was close to a zero correlation in the majority of the other measures of narrative and executive functioning. When the correlations were run using the English narratives for the monolingual speakers only, the overall pattern of results were similar to results presented in Table 4, yet the correlations between English HPA and the DCCS errors and the TOL total move scores were larger in magnitude (*r*s = -.37 and .33 respectively).

Table 4. Pearson Correlations between Narrative Complexity and Executive Functioning Measures

English Narrative Complexity (*n* = 76) (controlling for age)	GNG Comm. Errors	HTKS Total Score	Simon Effect	DCCS Block 3 Errors	TOL Total Move Score	BRIEF Global Ex. Comp.
Number of Unique Words	.17	.16	.08	-.11	.18	-.08
NAP- Short Form	.05	.28*	.09	-.09	.03	.09
SG (Total # of Elements)	.03	.16	.01	-.18	.14	-.07
High Point Analysis	.04	.07	.18	$-.20^{+}$	.29**	-.13
Spanish Narrative Complexity (*n* = 35) (controlling for gender & receptive vocab TVIP)						
Number of Unique Words	-.05	.22	.10	-.23	.07	.00
NAP- Spanish	.05	.21	.21	-.19	.48**	.03
SG (Total # of Elements)	-.02	.19	.19	$-.31^{+}$	53**	-.10
High Point Analysis	-.11	.13	.21	$-.30^{+}$	.05	-.04

$^{+}$ $p < .10$, *$p < .05$, **$p <.01$

Spanish

A structural equation model was run in Mplus to determine if there were a relation between Spanish narrative complexity and executive functioning, however, the solution did not successfully converge, likely due to the smaller sample size. Therefore, Pearson correlations were also run among each narrative complexity score in Spanish and each executive functioning task score, controlling for gender and receptive vocabulary (i.e., TVIP; see Table 4). There were no significant correlations between the number of unique Spanish words and any EF task. For the NAP-Spanish and for story grammar, there were significant positive correlations with the TOL total move score, such that children who scored higher on the NAP-Spanish or produced more story grammar elements in a narrative also had better performance on the Tower of London task. For both story grammar and HPA, there were marginal negative correlations with the number of DCCS errors; in other words, children who produced more story grammar elements in a narrative or scored higher on the HPA made marginally fewer errors on the DCCS task. Pearson correlations were also run without controlling for gender and receptive vocabulary, and the results remained to be the same with one exception: there was a significant negative correlation between HPA and the number of DCCS errors. These results reveal that there was a similar pattern found in the English narrative by executive functioning correlations, such that the majority of these individual correlations were close to zero, while the DCCS and Tower of London tasks were the only executive functioning measures that were meaningfully correlated with Spanish narratives.

Question 4. How does the relationship between narrative complexity and executive functioning vary by language group/degree of bilingualism and/or age?

A multi-group moderation model using structural equation modeling in Mplus was run to test if the relation between narrative complexity (the IV) and executive functioning (DV) varied by language group (Moderator; monolingual, balanced bilingual, and DLL), however, the solution did not successfully converge. A series of moderated multiple regressions was, thus, run to test the potential moderating effects of language group, degree of bilingualism, and age on the relation between narrative complexity and executive functioning. To reduce the number of group comparisons, only two language groups (i.e., monolingual vs. all bilingual children) were used in the regression analyses. In Step 1, an executive functioning score (DV) was regressed onto a narrative complexity score (IV) and moderator (M). Language group was coded into a dichotomous code where monolinguals were treated as the reference group. An interaction term was calculated by multiplying the dichotomous language group code with narrative complexity. In Step 2, the interaction term of narrative complexity x language group was added to the model. If the change in r^2 is significant from Step 1 to Step 2, then there would be evidence that language group moderates the relation of narrative complexity and executive functioning. This was repeated with all combinations of English narrative complexity and executive functioning scores. Next, this process was then repeated using the interaction of narrative complexity and degree of bilingualism (as a continuous variable) and age in Step 2. Finally, the same process was repeated using Spanish narrative complexity and exploring the moderating effects of language group (DLLs vs. balanced bilinguals), degree of bilingualism, and age. Due to the multiple comparisons described above, a Bonferroni correction was used to reduce the chance of making a Type I error. More specifically, each of the six executive functioning measures was regressed onto each of the four narrative complexity measures (i.e., 24 regression analyses) for each of the three moderators (i.e., 24 x 3 = 72 regression analyses), and this entire process was repeated for the two languages (i.e., 72 x 2 = 144 total regression analyses).

English

Overall, there was little to no evidence for any moderating effects of language group, degree of bilingualism, or age on the relation between English narrative complexity and executive functioning, regardless of a more conservative alpha level. More specifically, the addition of the interaction term did not add a significant change in the amount of variance accounted for by the model at a *p*-level of .05 or a more conservative estimate in each regression model that was run. There was an overall pattern that was in the hypothesized direction, such that the standardized regression coefficients were in favor of bilingual children and older children having a stronger relation between each narrative measure and each executive functioning assessment, however these were far from significance likely due to small sample sizes. Results of the two-level language group moderator yielded two significant interactions. The first was a moderating effect between the number of unique English words and the Simon Effect score. In Step 1, the number of English words and the language group did not significant predict the Simon Effect score, $F(2, 71) = 1.13$, $p = .330$, $R^2 = .03$. However, in Step 2, the addition of the interaction term did yield a significant result to the overall model, $F(3, 70) = 2.69$, $p <$

..05, $\Delta R^2 = .07$. In Step 2, the coefficient for the interaction term was significant, $B = -1.20$, $SE = .50$, $t = -2.38$, $p < .05$. Upon visual inspection of the interaction and running univariate post-hoc analyses, for bilinguals, the number of unique English words did not predict performance on the Simon task, but for monolinguals, as the number of unique English words increases, the Simon Effect score also increases, indicating, somewhat confusingly, that monolingual English speakers who also produced more unique English words had poorer performance on the Simon task.

The second significant interaction was a moderating effect between the NAP-Short Form in English and the Head Toes Knees Shoulders score. In Step 1, the NAP-Short Form significantly predicted the HTKS score, $B = .23$, $SE = .08$, $t = 2.75$, $p < .01$, and the overall model was significant, $F(2, 71) = 4.41$, $p < .05$, $R^2 = .09$. In Step 2, the addition of the interaction term did yield a significant result to the overall model, $F(3, 70) = 4.56$, $p < .01$, $\Delta R^2 = .05$. In Step 2, the coefficient for the interaction term was significant, $B = .36$, $SE = .17$, $t = 1.28$, $p < .05$. Upon visual inspection of the interaction and running univariate post-hoc analyses, for monolinguals, as hypothesized, the NAP-Short Form did not predict performance on the HTKS task, but for bilinguals, as the score on the NAP-Short Form increases, the HTKS score also increases, indicating that bilingual speakers who had higher NAP scores in English had better performance on the HTKS task.

Spanish

Overall, there was no evidence for any moderating effects of language group (i.e., DLL vs. balanced bilingual), degree of bilingualism, or age on the relation between

Spanish narrative complexity and executive functioning. The pattern was similar to the English results, and the interaction terms did not explain additional variance at a liberal $p < .05$ or a more conservative alpha level.

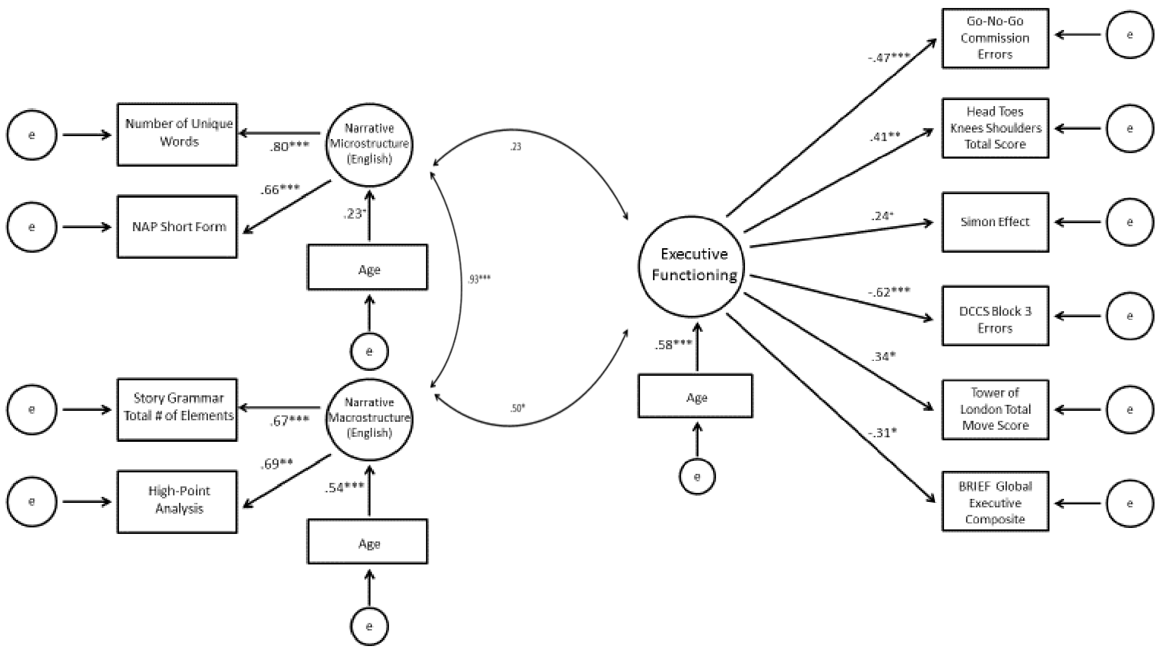
Question 5. In DLL and bilingual children, is the relation between narrative complexity and executive functioning the same for children whose parent-report of first language was English, Spanish, or both English and Spanish?

A series of Pearson correlations was run to see if and how the relation between narrative complexity and executive functioning varies by the parent-report of child's first language. Among the bilingual children, the breakdown was fairly equivalent across first language groups: $n = 12$ (33.3%) had English as their first language, $n = 11$ (30.6%) had Spanish as their language, and $n = 13$ (36.1%) had both English and Spanish. For the children whose parents reported that their "first language" was both English and Spanish, there was a variety of correlations between each narrative measure in both languages and the Go-No-Go number of errors, the Head Toes Knees Shoulders total score, the Simon effect, and the number of DCCS errors (the magnitude of *r*s between .53 and .76). For the children whose parents reported Spanish as the child's first language, the large correlations were primarily found in the Spanish narrative assessments and the Tower of London task (*r*s between .64 and .68), in addition to a large correlation between the Spanish HPA score and the number of DCCS errors ($r = -.75$). In other words, most of the strong relations between narrative complexity and executive functioning for Spanish as L1 children resided in the Spanish narrative assessments. For the children whose parents reported English as the child's first language, there were no strong correlations between any narrative measure and executive functioning measure. Therefore, there is some evidence that children whose first language was other than English had stronger relations between narrative complexity and executive functioning compared to children whose first language was English.

Question 6. Does the relation between narrative complexity and executive functioning depend on the level of analysis of narrative complexity (i.e., micro- vs. macro-level measures)?

English

A structural equation model was run to explore the different relations between micro-level and macro-levels of narrative complexity with executive functioning, but rather than creating one latent construct of narrative complexity as in Question 3, two latent constructs were created by using the number of unique words and the NAP short form score for the micro-level factor, and the total number of story grammar elements and the HPA score for the macro-level factor. The overall fit of the model was good, χ^2 (32) = 34.90, p = .332, and additional fit indices also revealed a good-fitting model, RMSEA = .04 and CFI = .97. A second model was run to include age as a predictor for both narrative complexity and executive functioning, as found from the results of Question 1 and the EFAs. The resulting model had very similar fit, χ^2 (39) = 43.88, p = .272, RMSEA = .04 and CFI = .96, but the parameter estimates were all similar in magnitude and level of significance.



Note. +*p* <.10, **p* < .05, ***p* <.01, ****p* < .001

Figure 5. Standardized Results of Relation between Different Levels of Narrative Complexity (in English) and Executive Functioning

As shown in Figure 5, there was a very strong positive correlation between micro- and macro-level narrative complexity scores (r = .93, p < .001). When isolating the relation between each level of analysis of narrative complexity, controlling for age, and executive functioning, the positive correlation for the micro-level construct was not significant (r = .23, p = .261), but the correlation was significant for the macro-level construct, as hypothesized (r = .50, p < .05). Using Steiger's (1980) method of testing the difference between two correlations, there was a marginal difference between the correlations between the micro- and macro-levels and executive functioning (z = 1.83, p = .067). In other words, the narrative construct related to organization and amount of important story elements was marginally related to executive functioning compared to the narrative construct related to the number of words and language-based grammar, which was in the direction of the hypothesis.

Pearson correlations were also run, controlling for age, between the English micro- and macro-level narrative variables and executive functioning. As shown in Table 5, the English micro-level variable was only correlated with the Head Toes Knees Shoulders Task ($r = .25$, $p < .05$). There was a marginal negative correlation between the English macro-level variable with the DCCS errors ($r = -.23$, $p < .10$) and a positive correlation with the TOL total move score ($r = .27$, $p < .05$). In other words, children who had organized, cohesive narratives that included important plot details also made fewer errors on the DCCS task and earned higher scores on the TOL task (i.e., better performance). These results indicated that the model run in Mplus was able to capture the relations among micro- and macro-level narrative complexity in a more sensitive way than running pairwise comparisons. Additional Pearson correlations between micro- and macro-structure and executive functioning separately for monolinguals and bilinguals because it was presumed that bilinguals would demonstrate better performance on the macro-structure of narrative discourse compared to monolinguals, however, there was no evidence of language group differences because the correlations were similar for both monolinguals and bilinguals.

Table 5. Pearson Correlations between Micro- and Macro-Levels of Narrative Complexity and Executive Functioning Measures

English Narrative Complexity (n = 76) (controlling for age)	GNG Comm. Errors	HTKS Total Score	Simon Effect	DCCS Block 3 Errors	TOL Total Move Score	BRIEF Global Ex. Comp.
Micro-Level Narr. Complexity[a]	.12	.25*	.10	-.12	.12	.01
Macro-Level Narr. Complexity[b]	.04	.14	.11	$-.23^{+}$	.27*	-.12
Spanish Narrative Complexity (n = 35) (controlling for gender & receptive vocab TVIP)						
Micro-Level Narr. Complexity[a]	.00	.23	.17	-.22	$.30^{+}$	.02
Macro-Level Narr. Complexity[b]	-.08	.20	.24	-.36*	.38*	-.09

[a]*Micro-Level was calculated by creating* z*-scores for the number of unique words and the NAP score, and then averaging the two* z*-scores*
[b]*Macro-Level was calculated by creating* z*-scores for the total sum of story grammar elements and the HPA score, and then averaging the two* z*-scores*
$^{+}$ $p < .10$, *$p < .05$, **$p < .01$

Spanish

A structural equation model using two latent constructs of Spanish narrative complexity was run in Mplus to test for differences in the strength of the relation between levels of narrative complexity and executive functioning, however, the solution did not successfully converge. Thus, in order to answer this question, first, each measure of narrative complexity (i.e., number of unique words, NAP, story grammar, HPA) was standardized using *z*-scores and then averaged accordingly to create an overall micro-level (i.e., number of unique words, NAP) and macro-level (i.e., story grammar, high-point analysis) score. Second, multiple regressions were run by regressing executive functioning scores (DV) onto the micro- and macro-level narrative complexity (IV). In

Step 1, based on the results of Question 1, gender and receptive vocabulary (i.e., TVIP) were included in the model. In Step 2, the average *z*-scores for the micro- and macro-levels of narrative complexity were added to the regression model. Standardized betas from Step 2 were then compared to see whether macro- or micro-level narrative scores were more associated with executive functioning, and this process was repeated for each measure of executive functioning. The results indicated that for all bilingual children, narrative complexity, regardless of level, did not predict any executive functioning measure in the regression analyses.

Pearson correlations were also run (controlling for gender and receptive vocabulary), between the Spanish micro- and macro-level narrative variables and executive functioning. As shown in Table 5, the Spanish micro-level variable was not correlated with any executive functioning measure, although there was a marginal positive correlation with the TOL total move score ($r = .30$, $p < .10$). Similar to the English macro-level results, the Spanish macro-level variable was correlated with two executive functioning measures; there was a negative correlation with the DCCS errors ($r = -.36$, $p < .05$) and a positive correlation with the TOL total move score ($r = .38$, $p < .05$). In other words, higher scores on measures of organizational complexity and details for a Spanish narrative were related to fewer errors on the DCCS task and higher scores on the TOL task (i.e., better performance).

DISCUSSION

Currently, researchers are beginning to use a narrative storytelling task to capture a child's expressive language in lieu of standardized assessments. Narrative storytelling tasks are easy to administer and, depending on the coding system, can be reliably coded without the need for transcribing the speech sample (Justice et al., 2010). Narratives are now being seen as ecologically valid and culturally unbiased language samples (Bedore et al., 2010; Fiestas & Peña, 2004). Their use with bilingual children is of particular interest because bilingual children have been shown to score lower on standardized assessments in receptive and expressive vocabulary compared to monolinguals (Pearson et al., 1993), which may due to standardized language assessments using total vocabulary within one language as an indicator of language ability rather than giving credit for knowledge of the concept in either language (Core et al., 2013). The ability to tell a good, well-organized, and cohesive story requires an interaction of linguistic, cognitive, and sociocultural abilities (Coelho et al., 1990; Silliman et al., 2002). Much of the research in narrative competence focuses on the linguistic and sociocultural aspects, with less emphasis on the examining how narratives relate to cognitive skills. Thus, the current study examined the narrative structure and complexity of stories during a storytelling task and related that to several direct measures of executive functioning for children in different monolingual and bilingual language groups. Additionally, potential moderators were tested to see if relations between narrative complexity and executive functioning were the same for monolingual and bilingual children, for children with varying degrees of bilingualism, and for children of different ages.

Prior research has not tested for a latent construct of overall narrative ability in a comprehensive manner, although one study by Terry et al. (2013) used several different measures of narrative complexity, including micro- and macro-level coding schemes, in order to explore how each scheme related to one another and how consistent the scores were over time. In the current study, a latent factor of English narrative complexity was identified, which included a combination of two micro- and two macro-level measures of complexity. A latent factor structure was also found in Spanish, and the parameter estimates were similar to those found in English. These results indicate that there is evidence for an underlying construct of narrative ability that is comprised on language-based measures (e.g., the number of unique words) and also macro-level organizational-based measures (e.g., plot builds up to a climax followed by a resolution), and that the same factor structure can be found in either English or Spanish in a group of young children with varying degrees of bilingualism. Terry et al. (2013) also found strong relations among a variety of narrative measures which suggest that children who use a wider variety of words also used more complex grammar and more elaborately organized narratives, but had not directly tested for a latent construct of narrative complexity.

Predictors of Narrative Complexity

Age, gender, and receptive vocabulary were tested as predictors of narrative complexity. For English narrative complexity (the primary, dominant language for most of the sample), age was found to be a significant predictor, such that as the age of the child in the study increased, the stories were coded as more complex and elaborate on the latent narrative factor. Individual correlations between each narrative measure and age in months revealed that the two macro-level measures were significantly related to age while the two micro-level measures were not. Compared to the age range of the current study, studies of narrative involving children between ages 3 and 5 often find age-related differences in the more language-based coding systems of grammar and vocabulary (Curenton & Justice, 2004; Muñoz et al. 2003). It is possible that number of unique words and grammatical features such as adverbs and prepositional phrases are not sensitive enough measures to detect differences among older children, such as those used in the present study. Similar to this study, other studies that include coding of macro organizational and referential features have also found age-related differences in children between ages 4 and 8 (Gutierrez-Clellen & Heinrichs-Ramos, 1993; Gutierrez-Clellen & Hofstetter, 1994).

Interestingly, age was not found to be a significant predictor for overall Spanish narrative complexity, although there was a strong correlation between age and the number of story grammar elements produced during the storytelling task. Prior studies that have found age-related differences in Spanish narratives included Spanish-speaking children with limited English proficiency (Gutierrez-Clellen & Heinrichs-Ramos, 1993; Gutierrez-Clellen & Hofstetter, 1994), whereas in this study there were no age-related differences in Spanish narratives. It is possible that the English-dominant DLLs in this study have not had enough exposure to Spanish in order to see age-related differences, which would dampen the age-related effect in the bilingual subsample. Upon further examination of the individual correlations, there was a similar pattern as the English results, but it is possible that the correlations were not significant due to the decrease in sample size from the overall sample to the bilingual-only subsample of 35 children with Spanish skills strong enough to get story data.

Instead, gender and Spanish receptive vocabulary predicted Spanish narrative complexity, such that girls (compared to boys) and children with higher Spanish receptive vocabulary produced more complex narrative in general. The gender differences were not due to girls demonstrating higher scores on Spanish receptive vocabulary because girls and boys had very similar scores on the TVIP (*M*s = 102.79 and 102.65 respectively). Gender differences are not typically reported, although a study of English/Spanish DLLs did find that boys produced more complex narratives compared to girls, but this was not found for Spanish stories (Uchikoshi, 2005). Spanish receptive vocabulary in this study was very strongly, positively related to each of the four Spanish narrative measures (*r*s between .34 and .71). This finding is similar to another study that has looked at the relation between narrative and receptive vocabulary in Spanish-dominant DLLs (Branum-Martin et al., 2009). The DLLs in the present study were largely English-dominant, so it is probable that children need good enough Spanish skills in order to tell a good story in Spanish. This distinction may explain why receptive language did not predict English narrative complexity because the English percentile scores on the PPVT were high across all children, regardless of language group (*M* %tile = 73.04, *SD* = 23.35). In other words, all of the children in the sample had good English skills, while the degree of Spanish skills varied widely within the bilingual subsample, largely due to the predominance of English-dominant DLLs.

English-Spanish Language Differences in Bilingual Children

Across all bilingual children in the sample, children produced more complex English narratives compared to Spanish narratives. When looking at the DLLs and balanced bilinguals separately, we find this remained to be true, although for the balanced bilinguals, the high point of the story (HPA) was found to be similar across languages. Among all bilingual children, there was a strong, positive correlation between the unique number of words produced in English and Spanish, and there was also a positive English/Spanish correlation the NAP balanced bilingual group and for the number of story grammar elements for the DLL group. The pattern from these results is that particularly for the balanced bilingual group, there is a strong cross-language association for the language-related coding measures, whereas for the DLLs, there is a cross-language association for the macro organization-related coding measures. This pattern was also found in prior studies of language differences in children's dominant and non-dominant languages, such that DLLs make more grammatical errors in their non-dominant language compared to their dominant language but could produce well-constructed narratives in their non-dominant language, whereas balanced bilinguals make a similar number of grammatical errors in both languages and show no differences in narrative quality (Bailey et al., 2008; Bedore et al., 2010; Cooperson et al., 2013; Fiestas & Peña, 2004; Gutiérrez-Clellen & Kreiter, 2003; Silliman et al., 2002). This pattern may have emerged in the English-dominant DLLs in this sample because the amount of variability was very high relative to the mean on all measures of Spanish narrative complexity, particularly in the language-based measures. A meta-analysis done by Branum-Martin and colleagues (2009) examined correlations between English and Spanish vocabulary measured in a variety of contexts. The results revealed the strongest positive relations between English and Spanish expressive vocabulary were found during a narrative storytelling task, as compared to receptive assessments, experimental measures, or the language proficiency battery of a cognitive assessment (Branum-Martin et al., 2009). The results from this study mimic this positive correlation, although their meta-analysis did not include cross-language comparisons for other measures of narrative complexity, so it is difficult to compare this result to prior studies.

Narrative Complexity Relating to Executive Functioning

Previous studies of narrative abilities in children with ADHD, pragmatic language impairment, and traumatic brain injury have suggested that there is an association between the organizational and cohesive aspects of narrative storytelling and executive functioning (Brookshire et al., 2000; Chapman et al., 1992; Ketelaars et al., 2012; Ygual Fernández et al., 2010), and a new study has encouraged further examination of this relation in bilingual and monolingual samples (De Houwer, Bornstein, Putnick, & Compernolle, 2015). Based on the findings from these studies, it was hypothesized that there would be a strong, positive correlation between overall narrative complexity and overall performance on executive functioning tasks. The latent factor of English narrative complexity was found to be positively correlated with a latent factor of executive functioning ($r = .49$, $p < .01$), but after controlling for age, the correlation was still there but only marginally significant ($r = .36$, $p = .06$).

This sizeable correlation is impressive, given that it was found with a low amount of power in the current study and controlling for age. The magnitude of the correlation between overall English narrative and overall executive functioning is similar to the magnitude of the standardized regression coefficients by regressing self-regulation and executive functioning measures onto early literacy skills, ranging from -.01 to .27 for inhibitory control and attention shifting in kindergarten while controlling for age and other background variables (Blair & Razza, 2007). Early narrative abilities have been found to be predict literacy and language skills (Snow, 1983), including intervention studies where children with whom mothers engage in narrative storytelling have shown improved language abilities up one year post-intervention (Peterson et al., 1999). Due to the evidence of concurrent relations between literacy and executive functioning and narrative and executive functioning, it would interesting to see how each of these domains develop over time. More specifically, the direction of the relation between narrative and executive functioning remains to be determined, and there are some researchers who believe narrative may simply be a manifestation of executive functioning (Coelho et al., 1990). However, I would presume that early narrative abilities would rely on language skills and the practice of specific cognitive abilities, which would in turn could impact later executive functioning skills.

When the micro- and macro-level measures were correlated separately with executive functioning controlling for age, the positive correlation between micro-level narrative and executive functioning was not significant (r = .23, p = .26), but it was for the macro-level (r = .50, p < .05). This finding supports the hypothesis that the organizational nature of narrative storytelling is more related to executive functioning than more language and grammar-based assessments. When age was taken out of this model, there was a significant difference between the micro- and macro-level correlations with executive functioning. Based on the results presented in Table 2, there were significant correlations between age and each of the macro-level measures (but not for the micro-level), so the amount of variance accounted for by age may have dampened the effect found in the structural equation model presented in Figure 5. In a study of children with pragmatic language impairment, there was a positive correlation between language production and executive functioning, but not for overall cohesion of the story (Ketelaars et al., 2012). Even though this result was opposite from findings from the current study, the effect went away when performance on a false-belief task—a measure of Theory of Mind (ToM)—was taken into account in the model, suggesting that ToM was accounting for the association between narrative language production and executive functioning. Colle, Baron-Cohen, Wheelwright, and van der Lely (2008) have suggested that narrative storytelling requires linguistic and social-cognitive skills, which are also key features of ToM. In their study of the association between ToM and narrative discourse, adults with autism spectrum disorders had less cohesive and more poorly organized narratives compared to typical controls, but these groups did not differ in the length of the story or number of utterances (Colle et al., 2008). Therefore, there appears to be support for the notion that there is something unique about the ability to organize a cohesive story, rather than focus on how long the story is or how many unique words were used. In order to tell a good story, the narrator must first create a mental representation of the story, organize it, and then identify the goals of the characters and interpret the problem-solving strategies that the character must use to obtain the goals (Skarakis-Doyle & Dempsey, 2008; McCabe, 1997). These skills are qualitatively different and emphasis the cognitive skills required to tell a story compared to using a narrative storytelling task to just obtain a more representative language sample of expressive language, which is largely reflective of linguistic abilities. Therefore, the language obtained from narrative storytelling is rich in both cognitive and linguistic skills on the part of the narrator. Although direction of effect is still unclear, parents and teachers may consider providing children with numerous opportunities for practicing their narrative storytelling, as improvements in narrative may also lead to improvements in executive functioning.

Narrative complexity in Spanish, however, did not yield a solution in SEM, and few correlations between individual narrative measures and executive functioning were found. One possibility was that the lack of significant findings with Spanish narratives was attributed to controlling for Spanish receptive vocabulary, although results were run with and without this covariate in the individual correlations with each measure of executive functioning, and the results remained very similar. There is a possibility that if the solution did converge in SEM, the results might have been similar to the English results due to the similarity in the individual correlation matrices between individual narrative measures and executive functioning measures in English and Spanish (see Table 4).

Moderating Effects of Narrative Complexity and Executive Functioning

It was hypothesized that if narratives were found to be related to executive functioning, bilinguals would demonstrate better performance on the macro-structure (i.e., organization and cohesion) of narrative discourse compared to monolinguals because bilinguals have been shown to have an advantage over monolinguals in a variety of executive functions, especially those that require planning and monitoring. In addition, in terms of age, the rate of development in executive functioning has been found to be non-linear; more specifically, studies have found a sharp increase in performance during the preschool years, followed by a steady increase (Zelazo et al., 2013; Zelazo & Carlson, 2012). There is support that there are increases in narrative competence over the preschool and early school age (Curenton & Justice, 2004; Gutierrez-Clellen & Heinrichs-Ramos, 1993; Gutierrez-Clellen & Hofstetter, 1994; Muñoz et al., 2003), but there is no evidence of a non-linear course of development. Therefore, it was hypothesized that the relation between narrative complexity and executive functioning would be stronger for older children, who may have experienced a sharp increase in executive functioning skills relative to their narrative competence. Attempts were made to identify how the relation between narrative complexity and executive functioning may vary by language group, degree of bilingualism, and age, but no moderating effects were found, in either English or Spanish. It is important to note that the lack of findings was not due to a conservative alpha level because of the large number of regression analyses; the moderation analyses revealed little evidence of any association. It is possible that the small number of children in each language group contributed to the lack of finding potential moderators. It is also possible that analyzing moderators individually for relations between each narrative measure and each executive functioning measure may have overlooked any interactions between the latent constructs of narrative and executive functioning, but unfortunately, solutions in a structural equation modeling platform were unsuccessful, likely due to the small sample size. There may not have been a moderating effect of age due to a similar rate of change in both narrative complexity and executive functioning across the sample. These moderations were exploratory in nature and were largely dependent on the outcome of overall association between narrative complexity and executive functioning. Future research, with larger samples including more Spanish-dominant DLLs is need to test the potential moderating effect of degree of bilingualism. It is important to keep in mind that even though the moderation analyses were not significant, there was still some suggestive evidence that early exposure to a non-dominant language may result in stronger relations between executive functioning and narrative skills as found in separately analyzing the correlations for each "first language" group. In other words, exposure to two first languages early on was related to stronger correlations between narrative and executive functioning. It is these children, who have used and switched back and forth between both languages from birth who are the bilinguals presumed to receive executive functioning benefits due to early experiences with bilingualism.

Limitations and Considerations for Future Research

The current study had several limitations that should be recognized. First, the sample size made language group comparisons challenging due to unequal and small *n*s per group, although these sample sizes are similar to other studies of the bilingual advantage for executive functioning (Carlson & Meltzoff, 2008; Martin-Rhee & Bialystok, 2008). Second, there were 10 children in the bilingual group (n = 2 balanced bilinguals, n = 8 DLLs) who refused to complete the narrative storytelling task in Spanish. In the original study, it was suggested by Hutchison (2012) that these children may experience pressure in their all-English public schools to only speak English in a school-like setting as it was when the children came into the university laboratory to participate in the study. Third, the children in the DLL group were English-dominant, even though the goal of the original study was to recruit Spanish-dominant DLLs. This was largely due to the limited locations to recruit bilingual children in the surrounding area from where the study took place. Many of the bilingual families who did participate placed an emphasis on English language skills because some were Anglo children in Spanish immersion school, which resulted in a sample of DLLs that are not representative of English Language Learners in the US. This limitation of English-dominant DLLs was also found in a similar study of the bilingual advantage in executive functioning that included DLLs (Carlson & Meltzoff, 2008). These English-dominant DLLs are still interesting to include in research, especially since there has been evidence that the DLLs in the original study did show improved performance on some of the EF tasks over the monolinguals, but not as large of an improvement compared to balanced bilinguals (Hutchison, 2012), suggesting that there can be a bilingual advantage in DLLs who only have limited exposure to a second language.

Use of Wordless Picture Books

A major advantage to using wordless picture books, as done in the present study, is that it allows for comparison of results across different studies. The results are even more comparable when researchers use the same *Frog* books across studies, and may be helpful in exploring how the narrative speech produced from an underrepresented sample of children (e.g., ethnic minorities, non-English speakers, children diagnosed with language impairment) may compare to other samples of typically developing children. However, it has also been suggested that asking a child to narrate a story based on a storybook is not a typical behavior for young children, especially for low-SES children (Morgan & Goldstein, 2004), which may be a novel experience and possibly be an invalid measure of narrative ability.

There are studies that have empirically compared the narratives of wordless picture books (specifically, the *Frog* books) to other narration methods. For example, McCabe et al. (2008) prompted 27 elementary school children to tell a personal narrative, in addition to narrating the wordless *Frog* picture book. The results of this study revealed that personal narratives were significantly more likely than chance to follow a classic narrative structure, such as describing a chronological sequence of multiple events and using past tense to refer to prior events, whereas narratives produced using the *Frog* books were more likely not to follow a narrative pattern, primarily using present tense to describe each picture in isolation (McCabe et al., 2008). A possible reason is that children frequently and naturally produce personal narratives that are scaffolded by their parents, more so than fictional narratives. Gorman et al. (2011) suggest that the content and construction of narratives are less constrained in personal narratives compared to using a wordless picture book, which limits the topic of interest. Nicolopoulou (1997) commented on the various narrative methods used in research, and described elicitation of spontaneous personal stories as being too messy to allow for rigorous measurement of comprehension, but also warned that using *Frog* books may limit the spontaneous character of the narrative by inhibiting children from inventing and creating their own stories. It is possible that the results would be different if natural stories had been used instead of those stemming from a wordless picture book, and future research should include both to test if there are different relations with executive functioning.

Challenges of Coding Narrative in Bilingual Samples

There are many challenges of coding narrative in bilingual children, particularly because bilinguals are not a heterogeneous group (Bialystok & Majumder, 1998). The amount of exposure to both languages varies widely from child to child, and there are important differences between balanced bilinguals and dual language learners to consider (e.g., socioeconomic status, immigrant status, language input in the home environment), that may account for differences on language assessments (Bialystok, 2001; Peña, & Halle, 2011). In bilingual language assessment, researchers must recognize the distribution of language skills within individuals because there may be variability on performance across different language measures (Jacobson & Walden, 2013). Researchers must also take into account the cultural assumptions of what should be emphasized in oral narratives, because practices may vary according to the culture (Gutiérrez-Clellen & Quinn, 1993). For example, the Hispanic culture emphasizes the description of personal relationships or reactions to events when telling a story, as opposed to emphasizing the causal sequence of events which is common in American culture (Fiestas & Peña, 2004).

Importantly, there is not enough normative data on narrative development for monolinguals, which may delay the progress to have normative data for bilinguals (Gutiérrez-Clellen & Iglesias, 1992). In general, there are not many language assessments that have been standardized or normed for bilingual children (Peña, & Halle, 2011); within narrative, coding systems in English have not typically been validated for bilinguals (Solari et al., 2013). Therefore, efforts should be made to develop measures that are robust enough to be valid in more than one language, yet sensitive enough to detect discrepancies between two languages. A concern with using HPA in Spanish-speakers has been raised by Bailey and colleagues (2010), such that the coding system may not capture the organization of a Spanish narrative, and the authors are wary of its use in research with non-English speakers. However, in this study, balanced bilinguals showed no difference in HPA scores in English and Spanish, so it is possible that this measure may not warrant concern for use in Spanish-speakers. Finally, the concerns listed have primarily been raised from research on the narratives of Spanish/English bilinguals in the United States, yet there are more concerns with the development of non-English coding systems in languages other than Spanish.

Future Directions

The current study is the first to examine the relation between narrative complexity and executive functioning in a comprehensive, latent factor model including typically developing children. Future research should apply this methodological approach to the preliminary research in this area with larger sample of DLL children as well as with children of atypical populations, such as children with ADHD or autism spectrum disorders since children in these populations have been found to have poor executive functioning skills (Russell, 1997). By testing these models in atypical populations, the factor structure can be compared against the model from this study, in particular to see if there are differences in the factor loadings for micro- or macro-levels of narrative complexity. It would also be interesting to see if these two constructs relate differently to executive functioning, based on evidence that children with mild/moderate traumatic brain injuries have deficits in both verbal memory and the organization of narratives, but do not have deficits in the length of narratives compared to typical controls (Chapman et al., 1992). Based on this prior research, it was hypothesized that bilingual children would demonstrate a stronger relation between narrative and executive functioning than monolingual children because bilingual children have been shown to have an advantage in several areas of executive functioning, which is directly in contrast to atypical populations with deficits in executive functioning (Russell, 1997). It would also be of interest to investigate the association between narrative storytelling and executive functioning in a longitudinal research design, in order to test the predictive properties of narrative storytelling in preschool-age children on executive functioning task performance upon entering school, and to also detect any age-related differences within the same children.

Conclusions

In conclusion, the current study was the first to test for an underlying latent construct of narrative complexity in both English and Spanish. Based on existing research with children from atypical populations, a relation between narrative complexity and executive functioning was tested in young children with varying degrees of bilingualism, and there was support for a relation but only in English. This study contributes to the larger body of narrative research that has only highlighted the cognitive skills required to tell a coherent and well-organized narrative without ever directly relating narratives to executive functioning. Since a relation has been established, future research can build off of the comprehensive models that were run in this study and apply them to other populations or in a longitudinal framework.

REFERENCES

Baddeley, A. D. (1986). *Working memory.* New York, NY: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science, 4*(11), 417–423.

Bailey, A. L., Moughamian, A. C., & Dingle, M. (2008). The contribution of Spanish-language narration to the assessment of early academic performance of Latino students. In A. McCabe, A. L. Bailey, & G. Melzi (Eds.), *Spanish-language narration and literacy: Culture, cognition, and emotion* (pp. 296–331). New York, NY: Cambridge University Press.

Baker, C. (2006). *Foundations of bilingual education and bilingualism* (4th ed.). Tonawanda, NY: Multilingual Matters.

Bamberg, M. G. W. (1987). *The acquisition of narratives: Learning to use language.* Berlin, Germany: Mouton de Gruyter.

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*(6), 498–510. doi:10.1016/j.jcomdis.2010.05.002

Berlin, L., & Bohlin, G. (2002). Response inhibition, hyperactivity, and conduct problems among preschool children. *Journal of Clinical Child and Adolescent Psychology*, *31*, 242-251.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70*, 636-644. doi:10.1111/1467-8624.00046

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, & cognition*. New York, NY: Cambridge University Press.

Bialystok, E. (2005). Consequences of bilingualism for cognitive development. In J. F. Kroll A. M. B. de Groot (Ed.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 417–432). New York, NY: Oxford University Press.

Bialystok, E. (2011). Coordination of executive functions in monolingual and bilingual children. *Journal of Experimental Child Psychology, 110*(3), 461–468. doi:10.1016/j.jecp.2011.05.005

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. doi:10.1037/0882-7974.19.2.290

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, *21*(6), 522–538. doi:10.1016/j.jneuroling.2007.07.001

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

Blair, C., & Razza, R. A. (2007). Relating effortful control, executive functioning, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647-663. doi: 10.1111/j.1467-8624.2007.01019.x

Branum-Martin, L., Mehta, P. D., Francis, D. J., Foorman, B. R., Cirino, P. T., Miller, J. F., & Iglesias, A. (2009). Pictures and words: Spanish and English vocabulary in classrooms. *Journal of Educational Psychology*, *101*(4), 897–911. doi:10.1037/a0015817

Brookshire, B. L., Chapman, S. B., Song, J., & Levin, H. S. (2000). Cognitive and linguistic correlates of children’s discourse after closed head injury: A three-year follow-up. *Journal of the International Neuropsychological Society*, *6*(7), 741–751. doi:10.1017/S1355617700677019

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, UK: Psychology Press.

Burns, F. A., de Villiers, P. A., Pearson, B. Z., & Champion, T. B. (2012). Dialect-neutral indices of narrative cohesion and evaluation. *Language, Speech, and Hearing Services in Schools, 43*(2), 132–152. doi:10.1044/0161-1461(2011/10-0101)

Cannizzaro, M. S., & Coelho, C. A. (2013). Analysis of narrative discourse structure as an ecologically relevant measure of executive function in adults. *Journal of Psycholinguistic Research, 42*(6), 527–549. doi:10.1007/s10936-012-9231-5

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*(2), 282–298. doi:10.1111/j.1467-7687.2008.00675.x

Chapman, S. B., Culhane, K. A., Levin, H. S., Harward, H., Mendelsohn, D., Ewing-Cobbs, L., … Bruce, D. (1992). Narrative discourse after closed head injury in children and adolescents. *Brain and Language*, *43*(1), 42–65. doi:10.1016/0093-934X(92)90020-F

Coelho, C. A., Liles, B. Z., & Duffy, R. J. (1990). Contextual influences on narrative discourse in normal young adults. *Journal of Psycholinguistic Research*, *19*(6), 405–420. doi:10.1007/BF01068887

Colle, L., Baron-Cohen, S., Wheelwright, S., & van der Lely, H. K. J. (2008). Narrative discourse in adults with high-functioning autism or Asperger syndrome. *Journal of Autism and Developmental Disorders, 38*(1), 28-40. doi:10.1007/s10803-007-0357-5

Cooperson, S. J., Bedore, L. M., & Peña, E. D. (2013). The relationship of phonological skills to language skills in Spanish–English-speaking bilingual children. *Clinical Linguistics & Phonetics*, *27*(5), 371–389. doi:10.3109/02699206.2013.782568

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in Spanish–English bilinguals from 22 to 30 months: Implications for assessment. *Journal of Speech, Language, and Hearing Research, 56*, 1637–1649.

Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70–89). New York, NY: Cambridge University Press.

Curenton, S. M., & Justice, L. M. (2004). African American and Caucasian preschoolers' use of decontextualized language: Literate language features in oral narratives. *Language, Speech, and Hearing Services in Schools*, *35*(3), 240-253.

Dart, S. N. (1992). Narrative style in the two languages of a bilingual child. *Journal of Child Language*, *19*(2), 367–387. doi:10.1017/S0305000900011454

De Houwer, A. Bornstein, M. H., Putnick, D. L., & Compernolle, C. (2015, March). *Narrative abilities in preschoolers: A bilingual-monolingual comparison*. Poster presented at the biennial meeting of the Society for Research in Child Development, Philadelphia, PA.

de Quirós, A. M. C., Lara-Alecio, R., Tong, F., & Irby, B. J. (2012). The effect of a structured story reading intervention, story retelling and higher order thinking for English language and literacy acquisition. *Journal of Research in Reading, 35*(1), 87–113.

Diamond, A. (2012). Activities and programs that improve children's executive functions. *Current Directions in Psychological Science, 21*(5), 335–341. doi:10.1177/0963721412453722

Dickinson, D. K., & Snow, C. E. (1987). Interrelationships among prereading and oral language skills in kindergartners from two social classes. *Early Childhood Research Quarterly, 2*(1), 1–25. doi:10.1016/0885-2006(87)90010-X

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test* (Rev. ed.). Circle Pines, MN: American Guidance Service.

Dunn, L. M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Circle Pines, MN: American Guidance Service.

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.

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. doi:10.1111/j.1467-9280.2008.02224.x

Espinosa, L. (2010). *Getting it right for young children from diverse backgrounds: Applying research to improve practice.* Upper Saddle River, NJ: Pearson.

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools, 35*, 155-168.

Frye, D., Zelazo, P., & Palfai, T. (1995). Theory of mind and rule-base reasoning. *Cognitive Development, 10*, 483-527. doi:10.1016/0885-2014(95)90024-1

Gándara, P., & Rumberger, R. W. (2009). Immigration, language, and education: How does language policy structure opportunity? *Teachers College Record, 111*(3), 750–782.

Gazella, J., & Stockman, I. J. (2003). Children's story retelling under different modality and task conditions: Implications for standardizing language sampling procedures. *American Journal of Speech-Language Pathology*, *12*(1), 61–72. doi:10.1044/1058-0360(2003/053)

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function*. Odessa, FL: Psychological Assessment Resources.

Goldstein, B. A., Bunta, F., Lange, J., Rodriguez, J., & Burrows, L. (2010). The effects of measures of language experience and language ability on segmental accuracy in bilingual children. *American Journal of Speech-Language Pathology*, *19*(3), 238–247. doi:10.1044/1058-0360(2010/08-0086)

Gorman, B. K., Bingham, G. E., Fiestas, C. E., & Terry, N. P. (2015). *Assessing the narrative abilities of Spanish speaking preschool children: A Spanish adaptation of the Narrative Assessment Protocol*. Manuscript in review.

Gorman, B. K., Fiestas, C. E., Peña, E. D., & Clark, M. R. (2011). Creative and stylistic devices employed by children during a storybook narrative task: A cross-cultural study. *Language, Speech, and Hearing Services in Schools*, *42*(2), 167–181. doi:10.1044/0161-1461(2010/10-0052)

Gosse, C. S., McGinty, A. S., Mashburn, A. J., Hoffman, L. M., & Pianta, R. C. (2014). The role of relational and instructional classroom supports in the language development of at-risk preschoolers. *Early Education and Development*, *25*(1), 110–133. doi:10.1080/10409289.2013.778567

Gutiérrez-Clellen, V. F., Calderón, J., & Weismer, S. E. (2004). Verbal working memory in bilingual children. *Journal of Speech, Language, and Hearing Research*, *47*(4), 863–876. doi:10.1044/1092-4388(2004/064)

Gutierrez-Clellen, V. F., & Heinrichs-Ramos, L. (1993). Referential cohesion in the narratives of Spanish-speaking children: A developmental study. *Journal of Speech & Hearing Research*, *36*(3), 559–567.

Gutierrez-Clellen, V. F., & Hofstetter, R. (1994). Syntactic complexity in Spanish narratives: A developmental study. *Journal of Speech & Hearing Research*, *37*(3), 645–654.

Gutiérrez-Clellen, V. F., & Iglesias, A. (1992). Causal coherence in the oral narratives of Spanish-speaking children. *Journal of Speech & Hearing Research*, *35*(2), 363–372.

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics*, *24*(2), 267–288. doi:10.1017/S0142716403000158

Heilmann, J., Miller, J. F., Iglesias, A., Fabiano-Smith, L., Nockerts, A., & Andriacchi, K. D. (2008). Narrative transcription accuracy and reliability in two languages.

Topics in Language Disorders, *28*(2), 178–188. doi:10.1097/01.TLD.0000318937.39301.76

Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology*, *19*(2), 154–166. doi:10.1044/1058-0360(2009/08-0024)

Hohle (2009). Crosslinguistic perspectives on segmentation and categorization in early language acquisition. In E.L. Bavin (Ed.), *The Cambridge handbook of child language* (pp. 125-144). New York, NY: Cambridge University Press.

Hu, L.-T., & Bentler, P. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling. Concepts, issues, and applications* (pp. 76-99). London, UK: Sage.

Hutchison, L. A. (2012). *Relations between executive functioning, second language fluency, and externalizing behavior problems in early childhood* (Unpublished doctoral dissertation). George Mason University, Fairfax, VA.

Iluz-Cohen, P., & Armon-Lotem, S. (2013). Language proficiency and executive control in bilingual children. *Bilingualism: Language and Cognition*, *16*(4), 884–899. doi:10.1017/S1366728912000788

Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition*, *15*(1), 58–74. doi:10.1017/S1366728911000538

Jacobson, P. F., & Walden, P. R. (2013). Lexical diversity and omission errors as predictors of language ability in the narratives of sequential Spanish–English bilinguals: A cross-language comparison. *American Journal of Speech-Language Pathology*, *22*(3), 554–565. doi:10.1044/1058-0360(2013/11-0055)

Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., & Gillam, R. B. (2006). The Index of Narrative Microstructure: A clinical tool for analyzing school-age children's narrative performances. *American Journal of Speech-Language Pathology*, *15*(2), 177–191. doi:10.1044/1058-0360(2006/017)

Justice, L. M., Bowles, R., Pence, K., & Gosse, C. (2010). A scalable tool for assessing children's language abilities within a narrative context: The NAP (Narrative Assessment Protocol). *Early Childhood Research Quarterly*, *25*(2), 218–234. doi:10.1016/j.ecresq.2009.11.002

Justice, L. M., McGinty, A. S., Zucker, T., Cabell, S. Q., & Piasta, S. B. (2013). Bi-directional dynamics underlie the complexity of talk in teacher–child play-based conversations in classrooms serving at-risk pupils. *Early Childhood Research Quarterly, 28*(3), 496–508. doi:10.1016/j.ecresq.2013.02.005

Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence*. Hoboken, NJ: John Wiley & Sons.

Ketelaars, M. P., Jansonius, K., Cuperus, J., & Verhoeven, L. (2012). Narrative competence and underlying mechanisms in children with pragmatic language impairment. *Applied Psycholinguistics, 33*(2), 281-303. doi: 10.1017/S014271641100035X

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. doi: 10.1186/1471-244X-5-40

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.

Lofranco, L. A. L., Peña, E. D., & Bedore, L. M. (2006). English language narratives of Filipino children. *Language, Speech, and Hearing Services in Schools*, *37*(1), 28–38. doi:10.1044/0161-1461(2006/004)

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.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: Transcription format and programs* (3rd ed., Vol. 1). Mahwah, NJ: Erlbaum.

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. doi:10.1037/a0014240

Mayer, M. (1969). *Frog, where are you?* New York, NY: Dial Books.

Mayer, M. (1974). *Frog goes to dinner.* New York, NY: Dial Books.

McCabe, A. (1997). Developmental and cross-cultural aspects of children's narration. In M. G. W. Bamberg (Ed.), *Narrative development: Six approaches* (pp. 137–174). Mahwah, NJ: Erlbaum.

McCabe, A., Bliss, L., Barra, G., & Bennett, M. (2008). Comparison of personal versus fictional narratives of children with language impairment. *American Journal of Speech-Language Pathology*, *17*(2), 194–206. doi:10.1044/1058-0360(2008/019)

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 & Practice*, *21*(1), 30–43. doi:10.1111/j.1540-5826.2006.00205.x

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*(4), 332–342. doi:10.1044/0161-1461(2003/027)

Nicolopoulou, A. (1997). Children and narratives: Toward an interpretive and sociocultural approach. In M. G. W. Bamberg (Ed.), *Narrative development: Six approaches* (pp. 179–215). Mahwah, NJ: Erlbaum.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, *43*(1), 93–120. doi:10.1111/j.1467-1770.1993.tb00174.x

Peña, E. D., & Halle, T. G. (2011). Assessing preschool Dual Language Learners: Traveling a multiforked road. *Child Development Perspectives*, *5*(1), 28-32. doi: 10.1111/j.1750-8606.2010.00143.x

Peterson, C., Jesso, B., & McCabe, A. (1999). Encouraging narratives in preschoolers: An intervention study. *Journal of Child Language*, *26*(1), 49–67. doi:10.1017/S0305000998003651

Poarch, G. J., & Van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113*(4), 535–551. doi:10.1016/j.jecp.2012.06.013

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*(2), 141-158. doi:10.1016/j.ecresq.2007.01.004

Poulin-Dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effects of bilingualism on toddlers' executive functioning. *Journal of Experimental Child Psychology*, *108*(3), 567–579. doi:10.1016/j.jecp.2010.10.009

Rumberger, R., & Tran, L. (2010). State language policies, school language practices, and the English learner achievement gap. In P. Gándara & M. Hopkins (Eds.), *Forbidden language. English learners and restrictive language policies* (pp. 86-101). New York, NY: Teachers College Press.

Russell, J. (1997). *Autism as an executive disorder*. New York, NY: Oxford University Press.

Schachter, R. E., & Craig, H. K. (2013). Students' production of narrative and AAE features during an emergent literacy task. *Language, Speech, and Hearing Services in Schools*, *44*(3), 227–238. doi:10.1044/0161-1461(2013/12-0034)

Schick, A., & Melzi, G. (2010). The development of children's oral narratives across contexts. *Early Education and Development*, *21*(3), 293–317. doi:10.1080/10409281003680578

Shallice, T. (1982). Specific impairments in planning. *Philosophical Transactions of the Royal Society of London B, 298*, 199-209.

Silliman, E. R., Bahr, R. H., Brea, M. R., Hnath-Chisolm, T., & Mahecha, N. R. (2002). Spanish and English proficiency in the linguistic encoding of mental states in narrative retellings. *Linguistics and Education*, *13*(2), 199–234. doi:10.1016/S0898-5898(01)00062-6

Silva, M., Strasser, K., & Cain, K. (2014). Early narrative skills in Chilean preschool: Questions scaffold the production of coherent narratives. *Early Childhood Research Quarterly*, *29*(2), 205–213. doi:10.1016/j.ecresq.2014.02.002

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, The Netherlands: North-Holland.

Skarakis-Doyle, E., & Dempsey, L. (2008). Assessing story comprehension in preschool children. *Topics in Language Disorders*, *28*(2), 131–148. doi:10.1097/01.TLD.0000318934.54548.7f

Snow, C. E. (1983). Literacy and language: Relationships during the preschool years. *Harvard Educational Review*, *53*(2), 165–189.

Solari, E. J., Zucker, T. A., Aghara, R., & Petscher, Y. (2013). *Testing and cross-validation of a modified version of the Narrative Assessment Protocol in a Spanish-dominant ELL pre-kindergarten sample*. Manuscript in review.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245-251.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing* (Vol. 2, pp. 53-120). Norwood, NJ: Ablex.

Terry, N. P., Mills, M. T., Bingham, G. E., Mansour, S., & Marencin, N. (2013). Oral narrative performance of African American prekindergartners who speak nonmainstream American English. *Language, Speech, and Hearing Services in Schools*, *44*(3), 291–305. doi:10.1044/0161-1461(2013/12-0037)

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, *9*(1), 77–110. doi:10.1016/0010-0285(77)90005-6

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*(3), 225–236. doi:10.1044/0161-1461(2007/024)

Uchikoshi, Y. (2005). Narrative development in bilingual kindergarteners: Can Arthur help? *Developmental Psychology*, *41*(3), 464–478. doi:10.1037/0012-1649.41.3.464

Westby, C. (1991). Learning to talk, talking to learn: Oral literate language differences. In C. S. Simon (Ed.), *Communication skills and classroom success* (pp. 334-357). Eau Claire, WI: Thinking Publications.

Woodcock, R. W. (1991). *Woodcock Language Proficiency Battery-Revised.* Circle Pines, MN: American Guidance Service.

Woodcock, R. W., & Muñoz-Sandoval, F. (1995). *Woodcock-Muñoz Language Survey.* Circle Pines, MN: American Guidance Service.

Yang, S., Yang, H. J., & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition*, *14*(3), 412–422. doi:10.1017/S1366728910000611

Ygual Fernández, A., Roselló Miranda, B., & Miranda Casas, A. (2010). Funciones ejecutivas, comprensión de historias y coherencia narrativa en niños con trastorno por déficit de atención con hiperactividad. [Executive functions, comprehension of stories and narrative coherence in children with attention-deficit hyperactivity disorder.]. *Revista de Logopedia, Foniatría y Audiología*, *30*(3), 151–161. doi:10.1016/S0214-4603(10)70163-7

Zelazo, P. D. (2006). The dimensional change card sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*, 297-301.

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). National Institutes of Health Toolbox Cognition Battery (NIH Toolbox CB): Validation for children between 3 and 15 years: II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development*, *78*(4), 16–33. doi:10.1111/mono.12032

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, *6*(4), 354–360.

BIOGRAPHY

Danielle L. Mead graduated from Pascack Hills High School, Montvale, NJ, in 2003. She received her Bachelor of Science from University of Mary Washington in 2007. She was employed as a research assistant in Washington, DC hospitals and National Institute of Mental Health for three years, and received her Master of Arts in Psychology from George Mason University in 2013.