Predictors of Performance on an iPad-Based Reading Comprehension Intervention Among Spanish-Speaking Dual Language Learners at Risk for Reading Comprehension Delays

by

Ashley Adams

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2017 by the
Graduate Supervisory Committee

María Adelaida Restrepo, Chair
Arthur Glenberg
Carol Connor
Beate Peter

ARIZONA STATE UNIVERSITY

August 2017

ABSTRACT

Purpose: The purpose of this study was to evaluate the efficacy of the EMBRACE Spanish support intervention for at-risk dual language learners and to determine which verbal and nonverbal characteristics of students were related to benefit from the intervention. The first study examined oral language and reading characteristics and the second study examined motor characteristics in predicting the children's outcomes on a reading comprehension intervention.

Method: Fifty-six participants in 2nd-5th grade were randomly assigned to one of two groups: 1) Spanish-support intervention, or 2) Spanish-support control. Outcome measures included performance on comprehension questions related to intervention texts, questions on the final narrative and expository text without strategy instruction, and difference scores on alternate forms of the Gates-MacGinitie (GMRT-4, MacGinitie, MacGinitie, Maria, & Dreyer, 2002) reading comprehension subtest administered pre-post-intervention. Multi-level hierarchical linear models were used to account for nesting of question within child within classroom. Regression models were used to examine the power of motor predictors in predicting Spanish and English language performance.

Results: Results from study 1 indicated that the intervention was most effective for narrative (vs. expository) texts and easy (vs. more difficult) texts. Dual language learners (DLLs) with lower initial English reading comprehension abilities benefitted more from the intervention than those with stronger reading skills. Results from Study 2 indicated that oral fine motor abilities predicted Spanish (but not English) oral language abilities in the expected direction (i.e. faster performance associated with higher language scores). The speed of /pata/ productions predicted reading comprehension during the intervention, but not in the expected direction (i.e. slower speeds associated with higher accuracy). Manual fine motor performance on tapping tasks was not related to language or reading. Conclusions: The EMBRACE intervention has promise for use with at-risk DLLs. Future research should take care to match text difficulty with child skills so as to maximize benefit from the intervention. Oral fine motor abilities were related to language abilities in DLLs, but only for the native language. Slower oral fine motor performance predicted higher accuracy on intervention questions, suggesting that EMBRACE may be particularly effective for children with weak fine motor skills.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

One of the most frequently cited theories of reading comprehension is the Simple View of Reading (SVR) which proposes that reading comprehension is the product of word recognition skills and language comprehension skills (Gough & Tunmer, 1986; Hoover & Gough, 1990). According to this model, when either of these skills is weak, the resulting comprehension of written text will also be weak. Numerous studies have confirmed that each of these sets of skills contribute unique variance to reading comprehension outcomes across the school years and explain between 65% and 85% of variance in reading comprehension outcomes (e.g., Aaron, Joshi, & Williams, 1999; Catts, Hogan, & Fey, 2003; Hoover & Gough, 1990). In addition, this model makes clear predictions about the source of reading comprehension difficulties in developing readers.

Among the most common risk factors for difficulties in English reading comprehension are being from a low socioeconomic status home and being a dual language learner (DLL) in a language minority context. As of the most recent National Asessment of Educational Progress (NAEP) report (2015), 78% of Hispanic children tested below proficient in both 4th and 8th grades, twice as many as White non-Hispanic children. The Hispanic DLL population that lives in poverty is steadily growing (Calderón, 2007; Hernandez & Napierala, 2014), resulting in an increasing number of children who have both of these risk factors and are likely to struggle with English reading comprehension. Children who have difficulties with reading comprehension are more likely to repeat grades or drop out of school (National Center for Education Statistics 2013), often with devastating effects for the individual, the family, and the community.

Poverty as a Risk Factor

According to the U.S. Census Bureau (2015), 23.6% of Latino families are living in poverty, compared to 10.1% of non-Hispanic white families. According to the Pew Research Center (2014), this number is even higher among the 18 and under age group, with 32% of Latino children living in poverty. Children who live in poverty are also at high risk for difficulties with reading comprehension. Children in the lowest socioeconomic groups have been shown to have cognitive scores on average as much as 60% lower than middle and high SES peers (Lee & Burkam, 2002). The NAEP reports that children from high SES backgrounds are twice as likely to score at or above basic in reading as their peers in poverty (Perie, Moran, & Lutkus, 2005). This gap between socioeconomic groups in reading ability has been shown to grow at an accelerated rate between 4th and 12th grade as the language demands of the curriculum increase (Stanovich, West, & Harrison, 1995). Children living in poverty are also more likely to live in neighborhoods served by schools with fewer resources, making it difficult for educators to have the tools necessary to close this gap. According to Neuman (2008), "professionals in the field of reading have no answer to the question of what to do about healthy children who come from poor environments who fail to learn to read" (p.2).

DLL Status as a Risk Factor

Being a DLL in the United States is another factor that contributes to risk for reading comprehension difficulties. It should be noted that learning a second language is not a risk factor in itself. Rather, the combination of demographic, political, and socioeconomic realities that surround DLL status in the United States place these children at risk. In some states, as many as 86% of all DLLs test below basic in reading comprehension at the 4th grade level (U.S. Dept. of Education, 2011). Nationwide, 68% of DLLs read below basic in 4th grade and only 8% read proficiently (NCES, 2015). According to the Condition of Education report in 2017, of the nearly 5 million school-age DLLs nationwide, 77.1% of these have Spanish as their first language (US Department of Education, 2017). Spanish-speaking DLLs are entering kindergarten at a significant disadvantage and the gap between their performance and that of their monolingual peers in reading comprehension often continues throughout their academic years (Planty et al., 2008).

Children who are DLLs are at specific risk to struggle with reading comprehension in their second language during their elementary school years. Evidence has suggested that, despite decoding skills in the normal range, DLL children's comprehension begins to fall behind in the 3rd grade and the gap widens throughout the rest of the school years (Droop & Verhoeven, 2003; Mancilla-Martinez & Lesaux, 2010; Nakamoto, Lindsey & Manis, 2007). While studies show that DLLs are able to develop word recognition skills on par with their monolingual peers, they consistently lag behind in reading comprehension ability in the later grades, making this an important group to target for reading comprehension interventions (Gottardo & Mueller, 2009; Lindsey, Manis & Bailey, 2003).

Reading Intervention for DLLs at Risk for Reading Comprehension Delays

The majority of research surrounding interventions for DLLs at risk for reading disabilities has focused on adapting strategies that have been effective for monolingual children (Vaughn, Mathes, Linan-Thompson, & Francis, 2005). These studies have revealed that effective interventions for this population focus on the areas of code-based skills, oral language skills, and higher-level comprehension strategies (August & Shanahan, 2006). For example, explicit and intense intervention for phonological skills such as rhyme, blending and segmenting phonemes has been effective in reading development for monolingual children with primary language impairment (PLI) (Al Otaiba, Puranik, Zilkowksi, & Curran, 2009) and for DLLs at risk for reading impairment (Haager & Windmueller, 2001; Klingner, Artiles, & Barletta, 2006; Nag- Arulmani, Reddy, & Buckley, 2003). Al Otaiba and colleagues (2009) suggest that children with more severe language delays (or PLI) will likely need ongoing phonological awareness training delivered by a speech-language pathologist. Oral language strategies, especially focused on development of new vocabulary, have been shown to be effective in several interventions for DLLs at risk for reading difficulties (Linan-Thompson, Vaughn, Hickman-Davis, & Kouzekanani, 2003; Vaughn et al., 2006), although improvements have limited generalizability. Finally, explicit instruction in comprehension strategies such as identifying cognate vocabulary words, comprehension monitoring, and inferencing are also potentially important components of reading interventions for DLLs (Jimenez, 1997; Jimenez, Garcia, & Pearson, 1996; Klingner & Vaughn, 1996).

An important question in the development of language and literacy interventions for DLLs has been whether to provide support in the first language and which skills transfer from the first to the second language. Much of this research is based on Cummins' interdependence theory that second language skills and proficiency are dependent on first language skills (Cummins, 1979). Broadly, bilingual interventions for DLLs with language impairment in the areas of vocabulary and syntax have been found to be at least as effective as monolingual interventions for improving outcomes in a child's second language (Restrepo, Morgan, & Thompson, 2013; Thordardottir, Cloutier, Ménard, Peland-Blais, & Rvachew, 2015). In the area of reading comprehension, research suggests that higher-order comprehension and metalinguistic strategies do transfer from the first to the second language (Genesse, Geva, Dressler, & Kamil, 2006; Jiménez, García & Pearson, 1996; van Gelderen, Schoonen, Stoel, Glopper, & Hulstijn, 2007). Additionally, Durgungolu (2002) demonstrated the importance of the development of formal definitions and academic language in both the first and second language for reading comprehension.

There are several theories as to why DLLs struggle with reading comprehension. The majority of research has linked reading difficulties to deficits in second language proficiency that frequently accompany DLL status (Droop & Verhoeven, 2003; Nakamoto, Lindsey & Manis, 2007). However, researchers also theorize that this relatively lower oral language proficiency also creates a bottleneck in processing that prevents children from successfully deploying higher-level comprehension strategies (Swanson, Sáez, Gerber & Leafstedt, 2004). The simulation theory of reading comprehension based on embodied cognition (Barsalou, 1999; Gallese, 2007, 2008; Gallese and Lakoff, 2005; Glenberg and Robertson, 2000; Zwaan and Taylor, 2006) allows for an additional prediction about the difficulty DLLs face when reading in a second language. That is, that DLLs may struggle to understand what they read in a second language because they have had fewer experiences in a specific language environment that prevents them from grounding the meaning expressed in the text in their own personal experiences (Adams, 2016).

Theoretical Background and Previous Studies

Simulation theory is based on the theory of embodied cognition, which proposes that meaning is not an abstract, amodal entity that exists outside of the body. Rather, the body and how it experiences the environment are integral to the creation of meaning. It is through this bodily experience that meaning is grounded (Lakoff, 1987). In language comprehension, there are two levels of grounding. The first is the evidence that comprehension of speech sounds invokes articulator-specific activation of speech motor cortex in sites overlapping with areas necessary for production of the same phonemes (Pulvermüller, et al., 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004). According to Pulvermüller and colleagues (2006), the fibers that connect the human motor and auditory regions necessary for speech perception and production are more well-developed than in any other nonhuman primate. This provides evidence of a special link between these two areas unique to humans. Therefore, in the last decade, there has been a renewed interest in the connection between speech production and perception and the neural correlates of these.

The second layer of grounding occurs at the level of the meaning of language. The Indexical Hypothesis (Glenberg & Robertson, 1999, 2000), based on the broader theory of embodied cognition, claims that language and reading comprehension occur through a process of simulation, in which the brain uses systems responsible for processing motor, sensory, and emotional information to simulate the content of the language being understood. This concept is perhaps best understood with a concrete example. Based on Zwaan and Taylor's (2006) work, if one hears a sentence such as "She turned the volume up on her car radio", part of the representation of this sentence is the clockwise turning motion necessary to accomplish this action. Therefore, one would expect activation of the motor strip in the hand area during comprehension of this sentence. This process is known as indexing, wherein linguistic symbols are indexed to perceptual, motor, or emotional experiences stored in the brain as perceptual symbols generated through previous experience (e.g. turning up the volume in the car) (Barsalou, 1999; Glenberg, 2011).

Several studies have explored this phenomenon and found evidence of motor abilities directly impacting language comprehension. Glenberg and Kaschak (2002) asked participants to listen to sentences describing transfer of concrete objects (e.g. "She gave the plate to him") and abstract ideas (e.g. "He told you a story") and to judge plausibility of the sentence by pressing a button that required hand motion either towards or away from the body. The authors found that hearing a sentence that described motion in the opposite direction from the direction that the participant had to move his/her hand to respond created interference and slowed response time. On the other hand, hearing sentences that described motion compatible with the movement necessary for the participant's response were speeded. Hauk, Johnsrude, and Pulvermüller (2004) found that participants who read action verbs related to a specific part of the body such as "lick", "pick", and "kick", showed site-specific motor strip activation (i.e. tongue areas were activated when reading "lick", hand areas when reading "pick", and leg areas when reading "kick") using fMRI. More generally, semantic information may be stored in areas categorically related to the word's meaning (e.g. smell-related words in olfactory cortex, visually related words such as concrete nouns in visual cortex, and action words in motor cortex) (González et al., 2006; Martin & Chao, 2001; Preissl, Pulvermüller, Lutzenberger, & Birbaumer, 1995; Pulvermüller, Lutzenberger, & Preissl, 1999; Pulvermüller, Härle, & Hummel, 2001).

The extent to which a human body can interact with a concept may be an important factor to take into consideration when examining the link between motor and language skills. Suggate and Stoeger (2014) studied vocabulary acquisition in preschool-aged children in relation to their ability to perform fine motor tasks such as peg-moving, bead-threading, and block-turning. They found that the relationship of fine motor skills to general vocabulary was mediated by the extent to which the child's body could interact with the concept represented by the vocabulary word (e.g. "feather" and "sawing" would be considered easier to interact with while words like "tunnel" and "diamond" would be considered less so). Similarly, Maouene, Hidaka, and Smith (2008) showed that acquisition of verbs was closely linked to a child's ability to move, with verbs related to the mouth being acquired earlier than verbs related to the arms, followed by verbs with less obvious mappings to body parts. Taken together, this evidence from scientists investigating embodiment theory as it relates to language indicates a strong link between the motor system and language and reading comprehension.

Fine Motor Skills, Language, and Academic Success

From the earliest stages of life, motor and communication skills develop in tandem. Iverson (2010) claimed that specific motor behaviors such as walking and object play provide the opportunity to acquire and practice skills critical for the development of language. Similarly, early impairments in motor skills are often associated with later problems with language acquisition (Amiel Tison et al., 1996).

This relationship continues throughout childhood, with motor abilities being important predictors of academic success (Cameron et al., 2012; Roth, McCaul, & Barnes, 1993). Roth, McCaul, and Barnes (1993) showed that, after controlling for vocabulary, visual skills, and demographic variables, the fine motor subtest score on the Early Prevention of School failure screening was the strongest predictor of referral for special education and the second strongest predictor of kindergarten retention. McPhillips and Jordan-Black (2007) showed that children's motor skills in preschool as measured by the Movement ABC (M-ABC; Henderson & Sugden, 1992) significantly predicted their word reading performance in 3rd grade. Cameron and colleagues (2012) found that the ability to copy a design was a particularly important fine motor ability that predicted later academic outcomes in academic knowledge, letter-word ID, and passage comprehension from the Woodcock Johnson III Test of Achievement (Woodcock, McGrew, & Mather, 2001). In fact, the relationship is quite sensible when one considers that as much as 46% of a elementary student's day is spent engaged in some kind of fine motor activity (Marr, Cermak, Cohn, & Henderson, 2003), primarily pencil and paper tasks. One might predict, therefore, that children who have achieved a higher level of automaticity in fine motor skills may have more remaining cognitive resources to learn new concepts, such as phonics (Berger, 2010). Or, as discussed earlier, the fine motor skill may directly impact language by a) affecting production, b) affecting perception of phonemes, and c) providing a basis for motor simulation of meaning.

Co-occurring Speech, Language, and Motor Impairments

Several researchers have explored the relationship among speech, language, and motor skills in children identified as having an impairment in one or more of these areas.

Hill (2001) conducted a meta-analysis of 28 studies that evaluated the relationship between language and motor skills in children with a diagnosed impairment in one of the two areas. She found that, depending on the criteria used to establish an impairment and the kind of tasks given to the participants in each study, between 40-90% of children who had a diagnosed language impairment would also qualify as having a motor impairment (Cermak, Ward, & Ward, 1986; Hill, 1998; Hill & Bishop, 1998; Rintala, Pienimaki, Ahonen, & Cantell, 1997; Robinson, 1991). The expected prevalence of this type of motor disorder in the general population is 6% (DSM-IV-TR, APA, 2000), indicating that even the lowest incidence found in the studies that Hill (2001) reviewed is markedly higher among groups identified as having speech and/or language impairment. In fact, since 1990, the diagnosis of developmental coordination disorder (DCD) has been found to overlap with poor attentional (Tervo, Azuma, Fogas, & Fiechtner, 2002), reading (Kooistra, Crawford, Dewey, Cantell, & Kaplan, 2005) and speech and language skills (Hill, 1998) in children between 7 and 16 years old.

Some researchers suspect a common genetic cause of motor and language problems. Bishop (2002) conducted a study comparing twins (ages 7-13 years old) with speech only or concomitant speech and language disorders against healthy controls. She found that twins with speech and/or language disorders were significantly slower on a finger-tapping task than healthy controls. She also suggested that the genes that put a child at-risk for speech and language difficulties might also affect motor ability. Similarly, Peter and colleagues have shown that children with severe speech impairments show significantly slower performance on oral and manual fine motor tasks requiring sequentially alternating movement (Peter, Button, Stoel-Gammon, Chapman, & Raskind, 2013; Peter, Matsushita, & Raskind, 2012; Peter et al., 2016).

The overlap found between language and motor impairments can be viewed from at least two different perspectives. The first is that a 3rd factor impacts the development of both systems. Ullman and Pierpont (2005) suggest that specific language impairment is not, in fact, specific and results from abnormal development of the brain structures responsible for procedural memory. These structures are theorized to impact both language and motor abilities. Similarly, other researchers theorize that the 3rd factor causing the co-occurrence of deficits is global processing delays (Bishop, 1994; Kail, 1994; Leonard, McGregor, & Allen, 1992; Norbury, Bishop, & Briscoe, 2001). However, the second possibility is that motor systems are somehow necessary for language processing, which is exactly what is predicted by the theory of embodied cognition.

Motor Skills, Embodiment, and Language in DLLs

Exploring the relationship between motor and language skills and how language is embodied among DLLs is made even more complex by the existence of two languages with potentially different trajectories of development. As Peña and Halle (2011) explain, even the assessment of motor skills can be impacted by such factors as the language the instructions are given in and the degree to which the interaction style between experimenter and child is culturally appropriate. Research concerning how motor abilities impact second language acquisition has primarily been limited to phonology and accent (Moyer, 2004). There are no studies known to this author that directly explore the link between motor, first and second language, and reading abilities simultaneously among DLL children. Nip and Blumenfeld (2015) examined late second language learners of

Spanish and found that second language proficiency was related to speech motor variability, speed, duration, and range of movement, and that variability was greatest for more linguistically complex utterances in the second language. Ullman (2005) claims that, since late second language learning may be more dependent on declarative (rather than procedural) memory, the relationship between motor and language abilities may be stronger in the first language than in the second.

Grounding of language does occur in the second language, but may depend on proficiency. De Grauwe, Willems, Rueschemeyer, Lemhöfer, and Schriefers, (2014) measured increased activation of the motor and somatosensory regions when participants were listening to simple verbs in their second language, suggesting that the embodiment of action verbs is not unique to the first language. However, proficiency in the second language appears to impact the extent to which this embodiment effect is present (Bergen, Lau, Narayan, Stojanovich & Wheeler, 2010; Harris, Ayçiçegi & Gleason 2003). Vukovic and Shtyrov (2014) showed that brain activity in sensorimotor areas during passive reading of words differed qualitatively depending on whether words were written in the participant's first or second language. The majority of these studies have been conducted with adult language learners. More research into motor, language, and reading skills in DLLs, especially children, is necessary to further elucidate the relationship among these abilities.

As described in Adams (2016), one could consider the Bilingual Interactive Activation Plus Theory (BIA+, Dijkstra & van Heuven, 2002) from an embodied perspective to explain relationships between motor and language skills in DLL children. According to this theory, embodied neural networks of sensorimotor experience may have differential resting-level activations that link speech motor and semantic information according to factors such as proficiency, word frequency, and amount of sensorimotor experience with a given object, action, emotion, etc. That is, the more experiences a child has had with a certain concept (e.g. a farm) in a specific language and cultural environment, the more likely that child would be to have stronger links between speech motor and semantic information (Jared, Poh, & Paivio, 2013). It is possible that the development of these links between speech motor and first language abilities develop early (i.e. from birth), and that the parallel relationships between oral motor and second language abilities may only show up in simultaneous bilinguals or highly proficient learners (which the current sample was not).

***EMBRACE*: An Embodied Approach to Improving Reading Comprehension**

The process of simulation and indexing of motor and perceptual symbols to their linguistic counterparts may not occur as naturally when comprehending written language as when comprehending oral language because it is rare that the language read in the text matches the present physical environment (Glenberg, 2011). For example, it is quite unlikely that, while reading *Moby Dick*, the reader is on a ship with a dangerous whale in their line of sight. This is in contrast to oral language comprehension, when language does frequently refer to the immediate environment, especially for young children (e.g. a child hears "Give daddy a hug" while someone points to daddy and gestures the action of hugging).

EMBRACE is designed to teach children how to create multi-modal simulations while reading. It is based on the theory described above and capitalizes on incorporating action and movement into reading comprehension. The intervention consists of two stages, physical manipulation (PM) and imagined manipulation (IM). In the PM stage, children are asked to click and drag objects on the screen to demonstrate understanding of target sentences. In the IM stage, children are asked to simply imagine such movement, and are randomly asked to choose an image that most closely corresponds to what they imagined. It is precisely the creation of these simulations that is hypothesized to be a critical area of weakness in DLLs with weak language skills and why I posit that *EMBRACE* will be an effective intervention for this group (Adams, 2016). Furthermore, understanding is closely related to the body's abilities because at least part of understanding is being able to take appropriate action based on the information that is understood. Therefore, requiring the children to physically (and imaginarily) interact with the text can improve reading comprehension by supporting the children in taking appropriate actions (simulating) while reading short stories.

An earlier, web-based version of the *EMBRACE* system called *Moved by Reading* has been shown to be effective in a population with learning disability. Marley, Levin, and Glenberg (2007) investigated the effectiveness of teaching a simulation strategy using PM in third-grade Native American children with documented learning difficulties based on failure to make appropriate academic progress. The children were randomly assigned to three conditions. In the PM condition, the children manipulated physical toys after listening to the experimenter read a sentence; in the visual condition, the children heard the experimenter read the sentence and then watched him manipulate the toys; in the free study condition, the children listened to the experimenter and were instructed to think about each sentence in the pause following the sentence. Children in the PM and visual conditions outperformed children in the free study condition by greater than one standard deviation for cued recall and free recall of propositions, objects, and actions.

Moved by Reading has been used with a variety of text genres. Glenberg, Willford, Gibson, Goldberg and Zhu (2012) demonstrated that 3rd and 4th graders (50% of whom were Hispanic) using PM and IM on a computer screen demonstrated improved comprehension compared to a control group. Children read both narrative and expository math-related texts and were able to correctly solve significantly more math story problems after the intervention than their peers in the control group. Furthermore, there was no significant difference between performance on narrative vs. expository texts (i.e., the intervention was equally beneficial for both types of text).

EMBRACE may be useful for understanding concrete and abstract concepts. Connor and colleagues (2014) included an extension of Glenberg's work called *Enacted Reading Comprehension* in their "comprehension tools for teachers" project. This intervention involved having 3^{rd} and 4^{th} grade children use the motion of their hands and arms to simulate meaning of literal and abstract texts about opposing forces. This strategy was used with science texts about earthquakes, persuasive texts involving debate, and a multi-chapter novel containing intra- and inter-personal conflict. The results demonstrated significant post-test gains for all three types of texts compared to a pre-test of relevant content knowledge.

Adams, Restrepo, and Glenberg (under review) showed that *Moved by* Reading was effective with typically developing DLLs in Arizona. Sixty-one children were divided into four groups (Spanish-English intervention, Spanish-English control, English-only intervention and English-only control). The children in the Spanish-English groups previewed the relevant vocabulary for the upcoming stories and listened to instructions, in Spanish and English, on how to read the stories. Then, they practiced with one oral story in Spanish before continuing to read in English. The control group was asked to simply think carefully about target sentences within the stories, whereas the intervention group was asked to manipulate images on the screen to demonstrate comprehension of sentences within the text. The children in the English-only groups heard all instructions and practiced in English. The children in the intervention groups outperformed the control groups in both the PM and IM stages of the intervention. In the PM stage, this effect was especially strong and driven by the Spanish-support intervention group's advantage over the control. (See below for summary of effect sizes).

Table 1

Between Group Effect Sizes Among DLLs (Cohen's d*)*

Comparison	PM Stage	IM Stage	Transfer
English-only Intervention vs. English Control	0.53	0.63	0.11
Spanish-support Intervention vs. Spanish-support Control	0.85	0.72	0.37

Second, the intervention was effective in a group of Spanish monolinguals. Adams and colleagues (under review) administered the intervention to 21 Spanish monolingual children. A comparison control group (n=20) was asked to simply think about target sentences. After each story, children answered a series of comprehension questions. The children in the intervention group significantly outperformed controls in both stages of the intervention and in a test of strategy generalization to unfamiliar texts based on comprehension questions asked at the completion of each story (See below for summary of performance).

Table 2

Between Group Effect Sizes Among Spanish Monolinguals (Cohen's d*)*

Condition	PM Stage	IM Stage	Transfer
Intervention vs. Control	.73	.78	.96

The current *EMBRACE* intervention has previously been shown to improve comprehension of narrative texts among Spanish-English DLLs who have strong English skills (but are poor decoders) *or* who are good decoders (but have poor English skills) (Walker, Adams, Restrepo, Fialko, & Glenberg, 2017). The PM stage has been shown to be particularly effective in improving comprehension among DLLs (Adams, Restrepo, & Glenberg, under review).

Based on the results of the above studies, as well as a review of the theory underlying the intervention, I hypothesize that the *EMBRACE* intervention will be effective in improving reading comprehension among DLLs at risk for reading comprehension delays for three reasons: 1) This intervention focuses on the development of a higher-level comprehension strategy, one area that has been shown to be a candidate for cross-linguistic transfer, by facilitating the activation of the motor and sensory systems during reading. 2) The interactive manipulation that takes place in the intervention will provide support for the creation of simulations using both semantic and syntactic information from the text. 3) The intervention incorporates both English and Spanish to assist children that have Spanish as their first language in learning a new strategy for reading.

In summary, DLLs are a group particularly at risk of reading comprehension delays. Simulation theory based on embodied cognition allows for a new approach to designing interventions for this group. Scientists researching embodied theories of language comprehension have demonstrated the grounding of language in action, represented in sensorimotor networks in the brain. This pattern of grounding has been shown to differ based on second language proficiency. The relationship of motor skills to second language abilities in children is an area still relatively unexplored. Considering that fine motor skills have been shown to predict academic performance, and that DLLs in English-only educational environments are a group particularly at risk for academic difficulties, this is an important area of study. Based on the evidence from adults, a plausible hypothesis would be that the grounding of language should exist in both languages, but to a lesser degree in the later-acquired or less proficient language. Given that reading comprehension requires grounding of concepts that are frequently not in the immediate physical environment, children with weaker embodiment of concepts in their second language may be at even higher risk to struggle with reading comprehension in their second language. As such, the *EMBRACE* intervention is a potentially effective intervention for DLL children with weak language and/or reading skills.

The studies described here add to the current body of knowledge on this topic in several ways. The recruitment process was specifically aimed at DLL children who were identified as experiencing difficulty with reading comprehension. The research questions for Study 1 were: 1) Does EMBRACE Spanish-support improve reading comprehension more than a control group a) on PM and IM intervention texts, b) on transfer texts, and c) on the change score of the Gates pre- to post-tests? 2) What child-level factor or factors best predict benefit from the intervention?

The second study examines in depth the connection between motor and language abilities. One of the critical predictions of the Indexical Hypothesis is a link between motor and language abilities, and the *EMBRACE* intervention is based on the idea of facilitating activation of the motor system while reading. Additionally, fine motor abilities have been shown to be important predictors of language and academic outcomes, but this topic has not been extensively studied among DLLs. Therefore, the second study allows for a deeper examination of motor-language-reading relationships among DLLs in order to determine the presence, strength, and direction of any correlation and if these abilities are important for performance on the intervention. As such, the questions addressed in the second study were: 1) What are the relationships among motor, first and second language, and reading variables in DLL children in 2nd to 5th grade? 2) Do fine motor variables (oral and manual fine motor skills) predict performance on L1 and L2 language measures? 3) Do oral and manual fine motor variables predict performance on *EMBRACE* comprehension questions? 4) Are there differences in motor performance between children with language impairment and those with typical development? Although the sample was the same for both studies, the first study addressed the more clinical, applied questions relating to efficacy of the intervention. In contrast, the second study addressed the more theoretical questions behind the intervention research.

Method

Participants

A total of 65 participants (33 female and 32 male) were recruited from three elementary schools in one Title I public school district in the Phoenix Metro area. Title I schools are those that receive federal funding due to having a high proportion of children from low-income homes. The average percentage of children qualifying for free or reduced lunch at these schools was 87%. (Arizona Department of Education, 2017). Of these, 19 (33%) children were attending dual language instruction and 37 (66%) were attending structured English immersion (SEI) programs. In dual language classrooms, the children spent half their day learning in Spanish from one teacher, and then shifted classrooms to learn in English from another teacher. However, in the SEI classrooms, all instruction was provided in English and four hours per day were spent focusing on English language development. Participants were identified by working with each school's 2nd-5th grade dual language and structured English immersion (SEI) teachers and the school speech-language pathologist to target DLLs who they thought would benefit from additional support in reading comprehension. Therefore, the sample was particularly at-risk for reading comprehension difficulties, and according to parent and teacher questionnaires, all children had either a parent or teacher who expressed concern about literacy and/or oral language development.

Results of the baseline language and reading comprehension standardized testing confirmed that this sample of children had low language and reading skills, as seen in the summary Table 3. Four children were on a current Individualized Education Plan (IEP, a document containing a plan for how the school intends to meet the educational needs of a child who has qualified for special education) for language impairment, two children had an IEP for articulation only, two children had an IEP for fluency issues, two children had an IEP for reading disability, and 3 children were in the process of being evaluated for special education due to reading concerns. An additional 12 children were referred for further speech and language evaluation as a result of baseline testing performed during the current project in which children referred received a score of 75 or below on both the CELF-English and Spanish, in accordance with district policies.

Participant selection criteria. All children met the following criteria: 1) the children had to be attending 2nd-5th grade; 2) the children had enough proficiency in English and Spanish to complete the baseline testing, based on their ability to produce at least simple sentences in a story retell task; and 3) the children could decode English words at an early independent reading level based on a criterion of performance at or above 50% accuracy decoding kindergarten-1st grade word lists of the Qualitative Reading Inventory, 5th Edition (QRI-5, Leslie & Caldwell, 2010). A total of 9 children were excluded from the study: two children moved out of the district between baseline testing and the intervention, two students did not have enough Spanish proficiency to complete the story retell, and five students scored below the criterion on the decoding measure. The final sample included 17 second-graders, 15 third-graders, 21 fourth-graders and 3 fifth-graders for a total of 56 children with an average age of 9;1 (range 7;6-11;3). Within each classroom, children were randomly assigned to the intervention or control group, at a ratio of 2:1 (i.e. 2/3 of children were assigned to the intervention group, and 1/3 were assigned to the control group). This random assignment strategy was used for two reasons: 1) to make sure that both the intervention and control groups were represented in each classroom, and 2) to provide the intervention to the most children possible while still having an appropriate comparison group. This resulted in an intervention group with 37 children (23F, 14M) and a control group with 19 children (9F, 10M).

Measures

Language proficiency.

Frog stories and SALT transcription. Each child retold a wordless storybook. For this sample, all children retold *Frog on his Own (*Mayer, 1967) in Spanish and *A Boy, a Dog, a Frog, and a Friend (*Mayer, 1967*)* in English. The child looked at the pictures while listening to an oral telling of the story in either Spanish or English. Immediately after listening, the child retold the story in the same language. The retellings were then sent to the SALT transcription center to be transcribed using SALT conventions. Unintelligible, abandoned and interrupted utterances were excluded. Utterances coded as code switching at the utterance level were excluded, but utterances containing code switching at the word level were included for all calculations. From these transcriptions, mean length of utterance (MLU), number of different words (NDW) and an ungrammaticality index (UGI) were calculated. These measures allowed for a detailed, objective, and continuous measure of language proficiency in both Spanish and English. Miller and colleagues (2006) found that these measures significantly predicted reading ability within and across languages in Spanish-English DLLs.

Questionnaires.

Parent questionnaire. A home language survey was sent home to all parents who gave their consent for children to participate. This questionnaire tracked information about concern for language and literacy development, languages spoken in the home, amount of time each language is used, and the home literacy environment (i.e. # of books in the house, # of times per week the child reads at home). This data was used to determine the presence of parent concern about language and literacy development. See Appendix E for results of this questionnaire.

Table 3

Participant Characteristics

Variable	Intervention (*n*=37)	Control (*n*=19)
Age in Months	109.6 (10.9)	107.0 (12.1)
Grade	3.2 (0.9)	3.2 (1.0)
CELF-English	74.1 (13.9)	81.1 (11.3)
CELF-Spanish	79.0 (16.8)	81.1 (14.4)
QRI Decoding	73.0 (27.3)	74.2 (21.1)
GMRT-4 Reading Comprehension Extended Scale Score	425.9 (38.6)	427.2 (31.8)
MLU English	8.0 (1.2)	8.3 (1.5)
MLU Spanish	6.5 (1.3)	6.7 (1.9)
Number of Times Child Reads Per week	4.7 (1.7)	3.8 (1.7)
WNV Standard Score	102.5 (13.5)	99.9 (14.9)

Note. No significant differences were present on any demographic variables (all ps>.05)

Teacher questionnaire. A teacher questionnaire gathered information about the teacher's estimate of each child's language proficiency in English and Spanish and reading comprehension abilities based on performance on curriculum-based measures in this area. It also allowed teachers to express concern (either not concerned, somewhat concerned or very concerned) for hearing, speech development, oral language development, literacy development, motor skills, thinking skills and social skills. This data was used to determine the presence of teacher concern about language and literacy development. See Appendix F for results of this questionnaire.

Baseline reading measures.

Qualitative Reading Inventory – 5th edition. (QRI-5; Leslie & Caldwell, 2010). This criterion-referenced measure was used to assess children's word decoding abilities. The QRI-5 contains word lists by grade based on content of materials found in basal readers and content-area textbooks. Reliability is reported between 98-99% if the administrator is trained in the QRI-5 guidelines. All children decoded the word lists for kindergarten through 5th grade, a total of 120 words. The test was discontinued if children decoded 7 consecutive words incorrectly to avoid frustration. This score was used to obtain a continuous measure of decoding ability and to screen out children who could not decode. Children who scored at or below 50% accuracy decoding kindergarten-2nd grade word lists were excluded from the study. This criterion was chosen because it was the approximate level at which children needed to be able to decode to read the intervention stories in a reasonable amount of time.

Gates-MacGinitie Reading Tests, Fourth Edition. (GMRT-4, MacGinitie, MacGinitie, Maria, & Dreyer, 2002) is a national test of reading comprehension with strong psychometric properties. Test–retest reliability coefficients of the GMRT are 0.89 to 0.93 for second through fifth grade (MacGinitie et al., 2002). Children in each grade completed the comprehension subtest Form S of the appropriate grade level (e.g. 2nd graders completed Form S Level 2, 3rd graders completed Form S Level 3, and so on) prior to the intervention. At the conclusion of the intervention, all children completed the comprehension subtest Form T at the same grade level. Difference scores between winter and spring extended scaled scores were used as a measure of improvement in reading comprehension.

Independent reading levels 1 and 2. First and second grade levels are designed to assess a student's level of early independent reading ability. This is the level at which they demonstrate fluent decoding and good comprehension without benefit of any teacher intervention. Only the comprehension subtest was administered, which is a timed test lasting 35 minutes.

Mature reading levels 3-10/12. These levels specifically measure a student's overall reading ability in Grades 3 through 10/12, and are divided into two sections that measure Vocabulary knowledge and Comprehension. These levels take 55 minutes to administer. Only the comprehension subtest was administered, which is a timed test lasting 35 minutes. The test was group-administered in the same groups in which the intervention was later administered.

Language measures.

***Clinical Evaluation of Language Fundamentals-4 English*.** (CELF-4 English; Semel, Wiig, & Secord, 2003). The CELF-4 is a standardized English assessment intended to determine the presence of a language disorder. The core language index score is composed of the following subtests: concepts and following directions, recalling sentences, formulating sentences, and word structure (for children 5-8 years old only), or word classes (for children 9-21 only). Standard scores of 85-115 are considered in the normal range. However, lower cut scores have been suggested for DLLs, especially those in high poverty schools (Restrepo, Castilla, Barragan, Nieto, under review). Inter-examiner reliability ranges from .88 to .99 and the test has fair to good sensitivity and specificity according to the examiner's manual. However, there are documented issues in using this test in isolation for the diagnosis of language impairment in DLLs (Leaders Project, 2013; Dollaghan & Horner, 2011).

Clinical Evaluation of Language Fundamentals–4 Spanish. (CELF-4 Spanish, Semel, Wiig, & Secord, 2006). The CELF-4 is a standardized Spanish assessment intended to determine the presence of a language disorder. The core language index score is composed of the following subtests: *conceptos y siguiendo direcciones* (concepts and following directions), *recordando oraciones* (recalling sentences), *formulación de oraciones* (formulating sentences), and *estructura de palabras* (word structure; for children 5-8 years old only), or *clases de palabras* (word classes; for children 9-21 only). It was developed specifically for Spanish speakers living in the United States as a parallel test to the CELF-4, English. It is not a translation of the English edition of the CELF-4. Test items incorporate grammatical forms appropriate for Spanish speakers and themes familiar to Spanish speaking students. Again, scores of 85-115 are traditionally considered in the normal range but lower cut scores are likely necessary for this at-risk DLL population (Restrepo, Castilla, Barragan, Nieto, under review). Inter-rater reliability ranged from .52 to .93 according to the technical manual and the sensitivity ranged from unacceptable to good depending on the cutoff score used. Specificity was considered good. There are documented issues in using this test in isolation for the diagnosis of language disorders in DLLs (Leaders Project, 2014).

Motor measures.

Computer tapping task. Two computer tapping tasks were used as a measure of manual fine motor skills following the published protocols by Gualtieri and Johnson (2006). A computer program custom-designed with Labview software (National Instruments, Austin, TX) was used to record tapping intervals. In the first task, repetitive tapping, the researcher instructed the children to look at the screen and focus on a round gray circle. They were instructed to start tapping on the spacebar as fast as possible from the time the gray start cue button turned bright green to the moment it returned to the gray color, which spanned ten seconds. The delay in onset of the start cue button was randomized between 2 and 4 seconds, so the children could not learn to anticipate the start of the task. In the alternating tapping task, the experimenter instructed the children to use their index and middle fingers to tap on the left and right arrow keys in an alternating fashion after the gray start cue button turned bright green to the moment it returned to the gray color (10 seconds).

In both tasks, the experimenter first demonstrated a trial, and then the child completed a practice trial. Children completed the first trial of each task with their right hands and then switched to their left hands for the subsequent trial. The switching was done in an effort to decrease the chance of hand fatigue. A total of 20 trials were administered to each child: 10 in the repetitive condition (spacebar tapping) and 10 in the alternating condition (one finger on each left and right arrow key) for a total of 5 trials per hand, per condition. If the experimenter noticed that the child made an error (e.g. began pressing buttons on the keyboard other than the spacebar or arrow buttons), the trial was discarded and the child was instructed to repeat the trial. Data from this task was collected in LabView as the millisecond at which each tap occurred. The average time in milliseconds between tapping (inter-tap) intervals was subsequently calculated for each task. Later, outlier inter-tap interval values of greater than 3 deviations in either direction from the mean were excluded to control for anomalies in the data or pauses due to external circumstances. The mean inter-tap interval in milliseconds for each task was recalculated after exclusion of the outliers and used as independent variables in the subsequent analyses. This process was performed twice by two separate experimenters, and the resulting numbers were identical indicating excellent inter-rater reliability.

Diadochokinetic (DDK) rates. This assessment is used to measure speed and regularity of oral movement of articulators (Fletcher, 1972). The researchers first modeled, then instructed, the children to produce as many of the tokens as fast as possible. These consisted of the monosyllable (/pa/) and the disyllable (/pata/). In parallel to the tapping tasks previously described, the productions of /pa/ can be thought of as the repetitive DDK task, and the production of /pata/ as the alternating DKK task. The researchers followed the methods to conduct this assessment as established by Fletcher (1972). Researchers attempted to obtain 20 and 30 tokens respectively for monosyllables and disyllables to calculate average syllable duration through the use of PRAAT software (version 6.0.26; Boersma and Weenink, 2016). The first token in a single breath (and the last when applicable in children who were not able to get a sufficient number of tokens in one breath) was discarded before analyzing. Outlier values of greater than 3 standard deviations from the mean in either direction were excluded from further analysis to control for anomalies in the data or pauses due to external circumstances. Average syllable duration (low numbers indicate rapid syllable repetition rates) was used as the variable of interest in all analyses. All syllable durations were double-scored. Inter-rater reliability of the /pa/ durations (92.9%) and of the /pata/ durations (91.1%) was considered excellent.

Bilingual Articulation and Phonology App. (BAPA; Fernandes, Kester, Bauman and Prath, 2011) This measure is a standardized assessment of articulation abilities at the single word level, specifically designed for children who are monolingual or bilingual English and Spanish speakers; it was used in our study to measure fine speech motor skills in the children's native and second language. The application contains 49 words in Spanish and 62 words in English that take into consideration all the phonemes in the respective languages in initial, medial, and final word position. Pictures of real objects appeared on the iPad screen and the child said the names of the items. If the child did not know the name of an item, he/she was instructed to hit the speaker button to hear words first then imitate the production. Total number of errors was coded at the word level (i.e. any word that contained an error was counted wrong only once, even if multiple errors were committed in the same word). I did not distinguish between errors typical of accented speech and true articulation errors (i.e. non-developmentally appropriate errors).

Cognitive measure.

Wechsler Nonverbal Scale of Ability. (WNV; Wechsler & Naglieri, 2006). The WNV is a commonly used measure of nonverbal cognitive ability, designed for linguistically diverse populations. The 2-subtest brief version was used in the present study. For children 8 and younger, this consisted of the matrices and recognition subtests. For children 9 and older, this consisted of the matrices and spatial span subtests. The matrices subtest requires children to choose from a field of 4 drawings the one they feel best completes a pattern. The recognition subtest requires children to look at an image for 3 seconds, then choose the matching image from a field of 4. Finally, the spatial span subtest requires children to watch as the experimenter points to a series of boxes in a specific order and then the child points to the boxes either in the same order or in reverse order. Scaled scores on each subtest were used to calculate an overall standard score, which was used at the independent variable in subsequent analyses.

Intervention. *EMBRACE* is an automated intervention implemented on the Apple iPad that aims to improve reading comprehension. The *EMBRACE* intervention consisted of three stages. The physical manipulation (PM) stage required children to read stories and move corresponding items on the screen. For example, if the sentence were "Carbon dioxide from the blood cell goes into the lungs," the child would use his or her finger to drag the carbon dioxide on the screen from the blood cell into the lungs (See Appendix G for an example story). This action is thought to foster comprehension and demonstrate understanding of target sentences within the story (Glenberg, 2011). Next, in the second stage, imagined manipulation, children were required to simply imagine they were moving the items on the screen, in the same way they previously did in the PM stage. Finally, a transfer stage tested the child's ability to use the imagined manipulation strategy in an unfamiliar story. The intervention was designed to scaffold the children to eventually imagine that they were manipulating items on the screen, in order to encourage strategy generalization to novel content. The intervention consisted of four narrative stories and four informational texts, each consisting of 5-7 chapters. Comprehension was assessed using both explicit and inferential comprehension questions at the completion of each story. A more detailed description of the intervention is provided below.

System. The two conditions (Spanish-support control and Spanish-support intervention) were delivered to children using *EMBRACE*, an iPad application implemented in Objective-C. On each day of the intervention, children were assigned one of 8 stories to read, which were indicated by a green bookmark on the active story, while all other story icons were a lighter gray. The first story the child read began with a brief introduction to *EMBRACE*. Then, the *EMBRACE* app read aloud the first chapter in each story with recorded audio files. The child read subsequent chapters in English. Each chapter began with a preview of the key vocabulary from that chapter. The child was required to tap on each word in the list to hear it pronounced, to hear a definition, and, where possible, to see a picture corresponding to the word highlighted. Once the child completed this vocabulary section, he or she continued reading the story. Below, I describe the functionality within the chapters, ways in which *EMBRACE* supports simulation and dual language support, as well as describe a control version of *EMBRACE*.

Simulation. Icons within the iPad system represented each story and each chapter within the story. A green bookmark marked the active chapter while the chapters that had not yet been assigned appeared in a lighter, dimmed gray. Each chapter consisted of several pages. The images that corresponded to the story took up the majority of the screen and the text was presented in a box at the top right of the screen. Students tapped a “Next” button to move from sentence to sentence within a page. Upon tapping “Next”, the just-read sentence was dimmed and the to-be-read sentence was colored in a bold blue or black font, depending on whether it was an action sentence (blue), or a non-action sentence (black). For non-action sentences, no simulation was prompted -- children read a sentence, and then hit the “Next” button to move to the next sentence. See appendix G for a sample page.

For action sentences in the PM stage, children read the sentence, and then used the images on the screen to demonstrate the content of the just-read blue sentences. For example, they might have read the sentence, “Lola, the bunny, hopped upstairs” and then moved the image of the bunny to the image of the upstairs part of the house. Since this was an iPad application, children touched an image to select it, and dragged it over to the desired position. If the images were moved incorrectly, a warning beep noise played and the image snapped back to its original position, providing both visual and auditory feedback that an incorrect manipulation has been made. If a manipulation was performed incorrectly three consecutive times, a pop-up window would appear telling the child "Need help? ¿Nescesitas ayuda? Now the iPad will show you how to complete this step." When the child pressed ok, the relevant items were highlighted and the action necessary was animated on the screen. Some sentences required several movements; for example, the sentence “The farmer took the hay into the barn” first required the child to move the farmer to the hay. The application grouped the farmer and the hay, and then the child had to move the grouped farmer and hay objects into the barn (as though the farmer were carrying the hay). Once the actions related to a sentence were completed, objects were ungrouped as appropriate (i.e., the farmer was disconnected from the hay so that the child might move the farmer on a later sentence).

In the IM stage, including some blue sentences after which the child was asked to choose between two possible relationships between objects monitored imagination of movement. For example, in the above sentence when the farmer was moved to the hay, the child was asked to select between two menu options (represented as drawings in two circular bubble shapes on the screen), one where the farmer is holding the hay, and one where the farmer is standing on top of the hay. Once the child was done with the simulation required by the sentence, they tapped the “Next” button to continue to the next sentence.

Spanish support. Spanish support was provided through the application in several ways. At the beginning of each chapter, a vocabulary preview page appeared with the key vocabulary words in Spanish and English. The child clicked on each word in both Spanish and English to hear the corresponding pronunciation, definition, and see the corresponding image highlighted. While the child was reading each chapter, some important or difficult words in the text were underlined. If the child needed help on an underlined word, the child could click on the word to hear the Spanish and English pronunciation, and where relevant, see the corresponding image highlighted. The *EMBRACE* app read the first story to the child in Spanish, rather than in English. The point of reading the first chapter aloud was to provide the child with relatively easily accessible background information that might be useful in reading and comprehending the remaining chapters. Thus, reading the chapter aloud in the child’s native language might have better prepared the child for subsequent chapters. Finally, children listened to the initial instructions on how to use the application, how to navigate from page to page, how to access vocabulary help, and how to manipulate the images in Spanish.

Control version of *EMBRACE*. In the Spanish-support control version of *EMBRACE*, students still tapped Next to move from sentence to sentence, and the current sentence was still colored blue or black, depending on whether it was an action sentence or a non-action sentence. However, children in the control group were simply instructed to think carefully about the blue sentences. Children saw the same images as the intervention group; however, children could not manipulate the images related to the story. Children still received vocabulary help in both languages before each chapter, and could still tap on words to receive their pronunciation in both languages on demand.

Procedure

Baseline testing and qualifying procedure. The first author (a bilingual speech-language pathologist) and trained research assistants completed all baseline and language testing. The first author scored all standardized tests. Examiners administered the CELF-4 English and Spanish core language subtests as well as the frog story retells in a one-on-one environment at the child's school, with all English measures being administered in a single session and all Spanish measures administered in a separate session. This testing took place during the child's school day. The QRI decoding measure was given during the English testing session. Experimenters administered the GMRT-4 in December-January in small groups as scheduling allowed.

Intervention procedure. Participants who qualified for the intervention were randomly assigned to either the control or intervention group, with 2/3 (37) children being assigned to the intervention group and 1/3 (19) children being assigned to the control group. This strategy was used to allow the highest number of children to participate in the intervention, while still having an appropriately sized comparison group. Stratified randomization was used at the teacher level to ensure that all levels of treatment existed in each classroom. Since the intervention was fully automated, including administration of the comprehension questions, participants were administered the intervention in groups of 5-11 at a time, with the supervision of an RA.

Table 4

Language Use During the Intervention

Group	Vocabulary Introduction	PM Instructions	PM Practice Chapter	PM Texts	IM instructions	IM Texts	Transfer Test Story
Intervention	Spanish then English	Spanish then English (move items)	Spanish	English	Spanish then English (imagine movement)	English	English
Control	Spanish then English	Spanish then English (think about sentences)	Spanish	English	Spanish then English (think about sentences)	English	English

Children in both the intervention and the control group read the texts over four weeks in four, 30-minute sessions per week (see below for weekly schedule). Children read 8 multi-chapter stories, 4 narrative texts and 4 expository texts, with the final narrative and expository text being a test of transfer. These texts were ordered according to difficulty based on the Flesch-Kincaid grade level (i.e. children first read 3 narrative texts of increasing difficulty, followed by 3 expository texts of increasing difficulty). This allowed children to acquire and practice the strategy intended by the intervention before attempting to use the strategy with more difficult texts. The final narrative and expository text used as a test of strategy transfer were of medium difficulty based on the Flesch-Kincaid scores in relation to the other texts children read during the intervention.

Within each chapter children previewed relevant vocabulary, then were asked to move (or imagine moving) items on the iPad screen to demonstrate understanding of individual sentences in each text. Comprehension was measured by asking follow-up questions at the end of each story, requiring explicit recall and inferencing. The final narrative and expository texts were tests of transfer, in which children read the stories without any instruction to manipulate or imagine manipulating items on the screen. Again, the control group went through the same steps, with the only difference being that they were instructed to simply think carefully about the sentences for which the children in the intervention groups were asked to manipulate items on the screen. The dependent variable was performance on the comprehension questions that were asked at the completion of the transfer stories. In addition, the Gates-MacGinitie reading comprehension subtest was administered as both a baseline and post-test measure to track progress in reading comprehension with a standardized measure.

Table 5

Intervention Schedule

	Day 1	Day 2	Day 3	Day 4
Week 1	Narrative Text 1 Chapters 1-4 PM	Narrative Text 1 Chapters 5-7 IM	Narrative Text 2 Chapters 1-3 PM	Narrative Text 2 Chapters 4-6 IM
Week 2	Narrative Text 3 Chapters 1-3 PM	Narrative Text 3 Chapter 4-6 IM	Expository Text 1 Chapters 1-3 PM	Expository Text 1 Chapters 4-5 IM
Week 3	Expository Text 2 Chapters 1-3 PM	Expository Text 2 Chapters 4-7 IM	Expository Text 3 Chapters 1-3 PM	Expository Text 3 Chapters 4-6 IM
Week 4	Narrative Text 4 TRANSFER TEST	Expository Text 4 TRANSFER TEST	Gates-MacGinitie Post Test Form T	

Outcome measures.

Intervention and transfer texts. For the comprehension and transfer texts, a python computer programming script was used to extract the accuracy data from the iPad-generated log files. For each comprehension question, I considered only the first attempt at answering. I coded the answer as 1 if correct or 0 if incorrect. For ease of analysis, these accuracy scores were then collapsed into 14 accuracy scores per child. This consisted of a PM and IM accuracy score for each of the six intervention texts and a single accuracy for each of the 2 transfer texts.

GMRT-4 change score. For the purposes of this analysis, I calculated a change score for each child by subtracting the extended scale score on Form S prior to the intervention from the extended scale score on Form T after the intervention. I chose the extended scale score because it allows for a continuous measure of progress both within and between children (MacGinitie, et al., 2000).

Statistical Analyses

Analytic Strategy. A three-level hierarchical linear model was built using HLM software (Raudenbush & Bryk, 2002) to account for the nesting of question within child within classroom present in this data. There were a total of 10 classrooms with 3-9 children per classroom. Using the HLM strategy allowed for estimation of treatment effects while accounting for classroom-level variance. Factors unique to the question including difficulty (Flesch-Kincaide grade level), type of text (narrative or expository), and stage of the intervention (PM or IM) were entered at level 1. Child-level factors were entered at level 2. See below for the strategy on which variables were entered. Finally, the teacher-level factor of grade was entered at level 3. This 3-level model was justified by examining the variance components in the unconditional model

Data Treatment. Since the primary coefficients of interest are at level 2 and interactions between level 2 and level 1 variables, the decision was made to group-mean center all level 1continuous and dichotomous variables and all level 2 continuous variables. The level 2 dichotomous variable of treatment was left dummy-coded for ease of interpretability of coefficients with respect to the treatment effect. If the treatment variable had been mean-centered, the intercept would be the mean accuracy for the control group and the treatment coefficient produced by the model would represent the difference between the mean of the treatment group and the weighted grand mean of the control and treatment groups. This is not an educationally meaningful statistic. Rather, by leaving the treatment variable dummy-coded, the intercept would still be the mean accuracy of the control group, but the treatment coefficient now represents the mean increase in accuracy associated with being in the treatment group. A total of .7% of the intervention comprehension data was missing. This data was excluded when constructing the MDM file.

Child Level Factors. Child-level factors were entered into the model based on theoretical background using the SVR model. However, due to the many baseline measures collected in this study, several steps were completed to determine which measures should be chosen to enter into the model. For example, a total of four measures of oral language were collected in each language: Spanish CELF-4 core language score, Spanish MLU, Spanish NDW, Spanish ungrammaticality index; English CELF-4 core language score, English MLU, English NDW, and English ungrammaticality index. Due to the limited number of participants, all four variables in each language and their potential interactions could not be entered into the model.

An initial model predicting accuracy on intervention comprehension questions was run with the four English language predictors entered at the child level to determine which was the strongest predictor of accuracy (see Appendix A, Tables A1 and A2 for results). Based on the high correlations among these measures, entering more than one into the model was likely to result in issues with convergence due to multicollinearity. Therefore, CELF-English was chosen as the English oral language variable to be entered into the model at the child level because it was the only significant predictor of accuracy of the four measures and it was highly correlated with the other 3 English language measures. A similar approach was taken when examining the Spanish language variables. Again, CELF-Spanish was the strongest predictor of accuracy and was highly correlated with the other Spanish language variables, so it was chosen as the Spanish oral language variable to enter into the final model (see Appendix A, Tables A3 and A4 for results). A final note that accuracy during the intervention was virtually identical for children receiving English-only education (mean = 59.87, SD = 21.79) and those in dual language (Spanish and English) education (mean = 58.75, SD = 21.64), so this variable was not entered into the model.

Results - Study 1

Research Question 1a: Does EMBRACE Spanish-support improve reading comprehension more than a control group on PM and IM intervention texts?

Building the Model. Based on the SVR model, oral language and decoding are the primary predictors of reading comprehension. Of course, among dual language learners, the oral language construct is made more complex by the co-existence of two languages developing in tandem. Therefore, both English and Spanish oral language variables were first entered into the model, along with the QRI decoding score and the corresponding interactions with treatment. This resulted in the following model using full maximum likelihood estimation, hereafter referred to as Intervention Model 1:

Level-1 Model

$\text{ACCURACY}_{ijk} = \pi_{0jk} + \pi_{1jk}*(INTERVENTION\ STAGE_{ijk}) + \pi_{2jk}*(TEXT\ TYPE_{ijk}) + \pi_{3jk}*(DIFFICULTY_{ijk}) + e_{ijk}$

Note. Intervention Stage was coded as -.5 = PM, .5 = IM; Text Type was coded as -.5 = Narrative, .5 = Expository

Level-2 Model

$\pi_{0jk} = \beta_{00k} + \beta_{01k}*(TREATMENT_{jk}) + \beta_{02k}*(CELF\text{-}ENG_{jk}) + \beta_{03k}*(CELF\text{-}SPA_{jk}) + \beta_{04k}*(QRI\ DECODING_{jk}) + \beta_{05k}*(TREATMENT\ BY\ CELF\text{-}ENG_{jk}) + \beta_{06k}*(TREATMENT\ BY\ CELF\text{-}SPA_{jk}) + \beta_{07k}*(TREATMENT\ BY\ QRI\ DECODING_{jk}) + r_{0jk}$

$\pi_{1jk} = \beta_{10k} + \beta_{11k}*(TREATMENT_{jk})$

$\pi_{2jk} = \beta_{20k} + \beta_{21k}*(TREATMENT_{jk})$

$\pi_{3jk} = \beta_{30k} + \beta_{31k}*(TREATMENT_{jk})$

Level-3 Model

$\beta_{00k} = \gamma_{000} + \gamma_{001}(GRADE_k) + u_{00k}$

$\beta_{01k} = \gamma_{010}$

$\beta_{02k} = \gamma_{020}$

$\beta_{03k} = \gamma_{030}$

$\beta_{04k} = \gamma_{040}$

$\beta_{05k} = \gamma_{050}$

$\beta_{06k} = \gamma_{060}$

$\beta_{07k} = \gamma_{070}$

$\beta_{10k} = \gamma_{100}$

$\beta_{11k} = \gamma_{110}$

$\beta_{20k} = \gamma_{200}$
$\beta_{21k} = \gamma_{210}$
$\beta_{30k} = \gamma_{300}$
$\beta_{31k} = \gamma_{310}$

ACCURACY$_{ijk}$ is the accuracy for child j in classroom k on the ith portion of the intervention (e.g. the PM stage of the easiest narrative text). Accuracy is a function of the coefficients of the three text-level factors (πs) plus an error term (e_{ijk}). π_{0jk} is the fitted mean for a given participant, which is also a function of the child-level factors plus a residual term (r_{0jk}). β_{00k} is the fitted mean for the classroom of students along with the classroom-level factor of grade and the error term (u_{00k}). γ_{000} is the fitted mean (intercept) for the entire sample. γ_{001} is the fixed effect of grade. γ_{010} is the fixed effect of treatment condition. γ_{020} is the fixed effect of CELF-English. γ_{030} is the fixed effect of CELF-Spanish. γ_{040} is the fixed effect of QRI decoding. γ_{050} is the fixed effect of the treatment by CELF-English interaction. γ_{060} is the fixed effect of the treatment by CELF-Spanish interaction. γ_{070} is the fixed effect of the treatment by QRI decoding interaction. . γ_{100} is the fixed effect of the intervention stage. γ_{110} is the fixed effect of the treatment by intervention stage interaction. γ_{200} is the fixed effect of text type. γ_{210} is the fixed effect of the treatment by text type interaction. γ_{300} is the fixed effect of text difficulty. γ_{310} is the fixed effect of the treatment by text difficulty interaction. The model allows for the partitioning of variance between text factors, child factors, and classroom factors.

Subsequently, all non-significant interaction effects (ps>.10) were removed from the model for the sake of parsimony. These included the Treatment by CELF-English (γ_{050}), Treatment by CELF-Spanish (γ_{060}), and Treatment by Text Type interactions (γ_{210}). This produced the model results as follows, hereafter referred to as Intervention Model 2

(See Table 6). After examining the results of this model, it was determined that CELF-Spanish should be removed from the model due to lack of significance. Although the main effect of Intervention Stage (PMIM) was not significant, it was left in the model due to its interaction with the treatment effect for further examination. This resulted in Model 3, where coefficients and significance levels were virtually identical to those reported above due to the small contribution of CELF-SPA in explaining any of the variance in Accuracy. A series of chi-squared tests were performed to determine if removing parameters from the model had resulted in decrease of model fit. Results are reported in Table 7 below. See Appendix D, Tables D1-D3, for a summary of effect sizes.

Table 6

Results of Intervention Model 2

Fixed Effect	Coefficient	Standard Error	*t*-ratio	Approx *d.f.*	*p*-value
for INTERCPT 1, π_0					
for INTERCPT 2, β_{00}					
INTERCPT 3, γ_{000}	56.37	2.53	22.28	8	<.001
GRADE, γ_{001}	4.47	2.28	1.96	8	.085
for TREATMENT, β_{01}					
INTERCPT 3, γ_{010}	3.56	2.40	1.49	41	.145
for CELFENG, β_{02}					
INTERCPT 3, γ_{020}	0.27	0.11	2.52	41	.016
for CELFSPA, β_{03}					
INTERCPT 3, γ_{030}	-0.002	0.08	-0.03	41	.980
for QRI, β_{04}					
INTERCPT 3, γ_{040}	0.35	0.11	3.09	41	.004
for TREAT by QRI, β_{05}					
INTERCPT 3, γ_{050}	-0.21	0.11	-1.98	41	.054
for PMIM slope, π_1					
for INTERCPT 2, β_{10}					
INTERCPT 3, γ_{100}	0.55	1.71	.32	596	.747
for TREAT, β_{11}					
INTERCPT 3, γ_{110}	-3.64	2.11	-1.73	596	.085
for TEXTTYPE slope, π_2					
for INTERCPT 2, β_{20}					
INTERCPT 3, γ_{200}	-23.03	1.12	-20.47	596	<.001
for DIFFICULTY slope, π_3					
for INTERCPT 2, β_{30}					
INTERCPT 3, γ_{300}	-2.90	0.89	-3.29	596	.001
for TREAT, β_{31}					
INTERCPT 3, γ_{310}	-2.23	1.04	-2.14	596	.033

Note. All coefficients (γ_{ijk}) reported are level 3 fixed effects. Coefficients with format γ_{i00} are level 1 (text level) fixed effects. Coefficients with format γ_{0j0} are level 2 (child level) fixed effects. Coefficients with format γ_{00k} are level 3 (classroom level) fixed effects. Coefficients with format γ_{i10} are level 2 by level 1 interaction fixed effects.

Table 7

χ^2 Tests Comparing Model Fit

Model	Deviance	Number of Estimated Parameters	χ^2 Difference Statistic	*d.f.*	Significance
Unconditional	5936.15	4	N/A		N/A
Model 1	5399.4	18	536.70	14	<.001
Model 2	5400.12	15	0.72	3	>.500
Model 3	5400.13	14	0.006	1	>.500

Due to lack of decrease in model fit, while estimating fewer parameters, Model 3 was reserved as the final model for answering research question 1a. Children with higher decoding and higher CELF-English scores also had higher accuracy scores, as seen by the positive coefficients and significant main effects of these two variables. The main effect of Text Type (narrative or expository) showed that children were significantly more accurate on narrative texts than on expository texts, but the lack of interaction with treatment indicated that the relative difference between treatment and control was the same for both types of texts. While treatment was not significant overall (See Appendix B for graphs of raw scores on the 6 intervention stories in each stage), there were several interactions that were worth examining further to explore the treatment effect. The Treatment by Decoding interaction suggested that children who were less proficient decoders benefited more from the intervention than those who began the intervention as more proficient decoders (Figure 1 below). The Treatment by Intervention Stage interaction suggested that the PM stage provided more of an advantage in comprehension over the control group than did the IM stage (See Figure 2 below). Finally, the Treatment by Difficulty interaction showed that the treatment effect was greater for easier texts than for more difficult texts (See Figure 3 below).

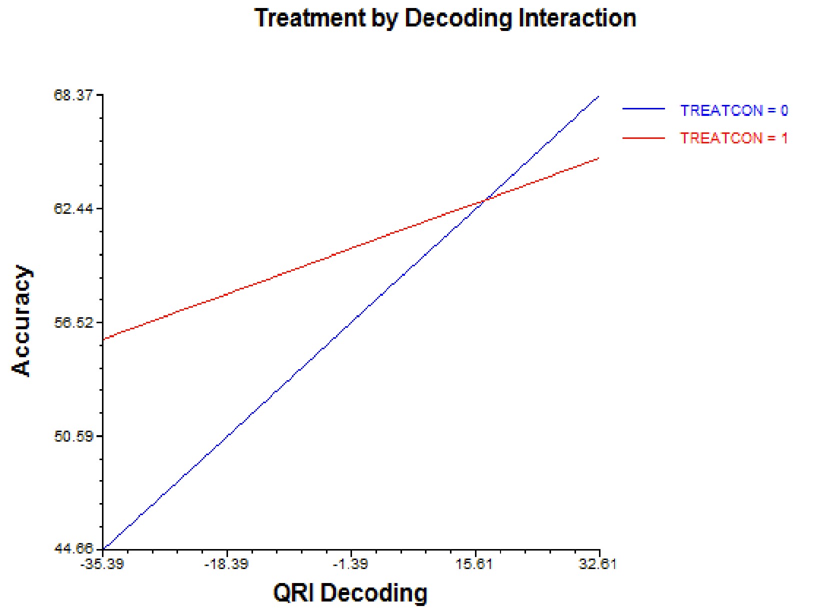

Figure 1. Treatment by decoding interaction.
Note. 0=control (blue); 1=intervention (red)

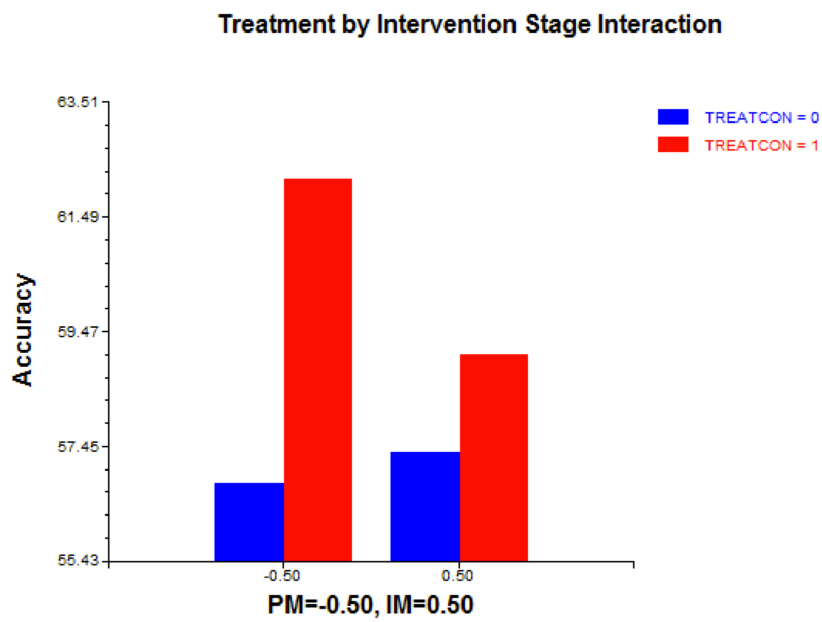

Figure 2. Treatment by intervention stage interaction.
Note. 0=control (blue); 1=intervention (red)

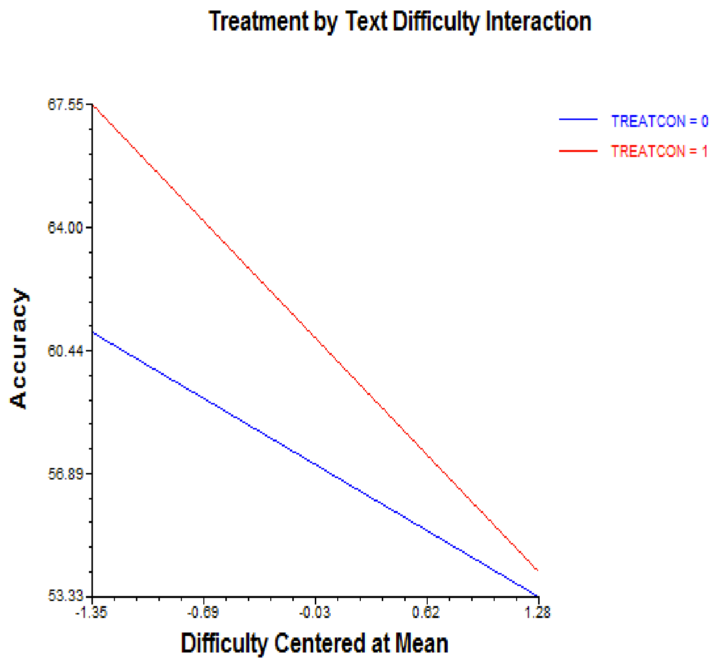

Figure 3. Treatment by text difficulty interaction.
Note. 0=control (blue); 1=intervention (red)

Research Question 1b: 1) Does EMBRACE Spanish-support improve reading comprehension more than a control group on transfer texts?

Building the Model. Model 1 was created in an identical fashion using the accuracy of the final two stories (one narrative and one expository) as the dependent variable. The only difference in the Level 1 model was that Text Type was the only variable entered, as Intervention Stage (PM or IM) was no longer relevant as children were not instructed to use any strategy during this stage and the images were static (i.e. no interaction was possible) during the reading of the transfer texts. Difficulty was redundant with Text Type as there were only two values for each variable. Similar to Model 1 in the previous analysis, the non-significant interactions were trimmed from

Transfer Model 1. The resulting model (Transfer Model 2) was retained as the final model and can be seen below in Table 8. See Appendix D, Tables D4-D6, for a summary of effect sizes. Model fit did not significantly decrease due to the removal of the non-significant interactions ($\chi^2(3)=2.29$, $p>.500$). As can be seen above, there were significant main effects of grade, CELF-English, and Text Type and a borderline significant effect for QRI decoding.

In contrast to the models of the intervention stages, there was a negative main effect of CELF-Spanish, such that for a one-point increase in CELF-Spanish score, the accuracy score decreased by .22%. There was an approximately 4% increase in overall accuracy for children in the treatment group, but this effect was not significant in the Transfer stage. While the Treatment by Decoding interaction was not significant in the final model, figure 4 below demonstrates a similar pattern as during the intervention; that is, children with poorer decoding benefitted more from the intervention than those with better decoding.

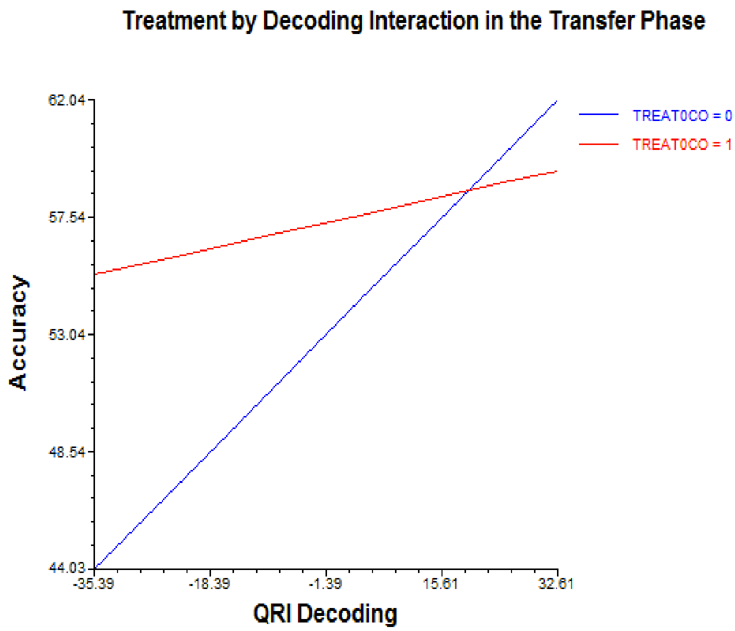

Figure 4. Treatment by Decoding Interaction in the Transfer Phase.
Note. 0=control (blue); 1=intervention (red)

Table 8

Results of Transfer Model 2

Fixed Effect	Coefficient	Standard Error	*t*-ratio	Approx. *d.f.*	*p*-value
for INTERCPT 1, π_0					
for INTERCPT 2, β_{00}					
INTERCPT 3, γ_{000}	52.46	2.44	21.50	8	<.001
GRADE, γ_{001}	7.80	1.95	4.00	8	.004
for TREATMENT, β_{01}					
INTERCPT 3, γ_{010}	4.02	3.06	1.31	41	.196
for CELFENG, β_{02}					
INTERCPT 3, γ_{020}	0.48	0.14	3.39	41	.002
for CELFSPA, β_{03}					
INTERCPT 3, γ_{030}	-0.22	0.10	-2.16	41	.037
for QRI, β_{04}					
INTERCPT 3, γ_{040}	0.26	0.14	1.92	41	.061
for TREAT by QRI, β_{05}					
INTERCPT 3, γ_{050}	-0.21	0.13	-1.65	41	.107
for TEXTTYPE slope, π_2					
for INTERCPT 2, β_{20}					
INTERCPT 3, γ_{200}	-20.00	1.167	-12.00	45	<.001

Note. γ_{000} is the fitted mean (intercept) of accuracy during the transfer stage for the sample. All coefficients (γ_{ijk}) reported are level 3 fixed effects. Coefficients with format γ_{i00} are level 1 (text level) fixed effects. Coefficients with format γ_{0j0} are level 2 (child level) fixed effects. Coefficients with format γ_{00k} are level 3 (classroom level) fixed effects.

Research Question 1c: Does EMBRACE Spanish-support improve reading comprehension more than a control group on the change score of the Gates-MacGinitie pre- to post-tests?

Since the Gates-MacGinitie change score was a single, unique score for each child, a two-level hierarchical model was constructed using restricted maximum likelihood estimation (the default for 2-level models). Examination of the variance components in the unconditional model suggested that 22.5% of the variance in Gates MacGinitie change scores was at the teacher level ($\chi^2(9)=27.59$, $p=.001$), justifying the decision to run a 2-level nested model. Similar to the previous two sets of analyses, the initial model was constructed with the CELF-English, CELF-Spanish, QRI decoding, and their interactions included at the child level. Grade was included at the teacher level. Non-significant interactions were trimmed and resulted in the following model. See Appendix D, Tables D7 and D8, for a summary of effect sizes.

Table 9

Results of Gates-MacGinitie change final model

Fixed Effect	Coefficient	Standard Error	*t*-ratio	Approx *d.f.*	*t*-value
for INTERCPT 1, β_{00}					
for INTERCPT 2, γ_{00}	9.59	5.03	1.91	8	.093
GRADE, γ_{01}	-9.94	3.62	-2.74	8	.025
for TREATMENT slope, β_1					
INTERCPT 2, γ_{10}	0.37	5.36	0.07	42	.946
for CELFENG slope, β_2					
INTERCPT 2, γ_{20}	-0.12	0.24	-0.50	42	.617
for CELFSPA slope, β_3					
INTERCPT 2, γ_{30}	-0.32	0.17	-1.86	42	.070
for QRI slope, β_4					
INTERCPT 2, γ_{40}	0.14	0.14	1.0	42	.325

Note. γ_{00} is the fitted mean (intercept) of Gates-MacGinitie change in the sample. γ_{01} is the fixed effect of grade. All coefficients (γ_{i0}) reported are level 1(child level) fixed effects.

The results indicate that neither treatment, CELF-English, nor decoding were significant in the model. Grade was significant and the coefficient was negative, suggesting that kids in 2nd grade had the highest Gates-MacGinitie change scores, but for each increasing grade, the Gates-MacGinitie change score decreased by 9.94 as can be seen below in Figure 5. CELF-Spanish was borderline significant, and the coefficient was negative, suggesting that for each one-point increase in CELF-Spanish score, the Gates-MacGinitie change score decreased by 0.32 points.

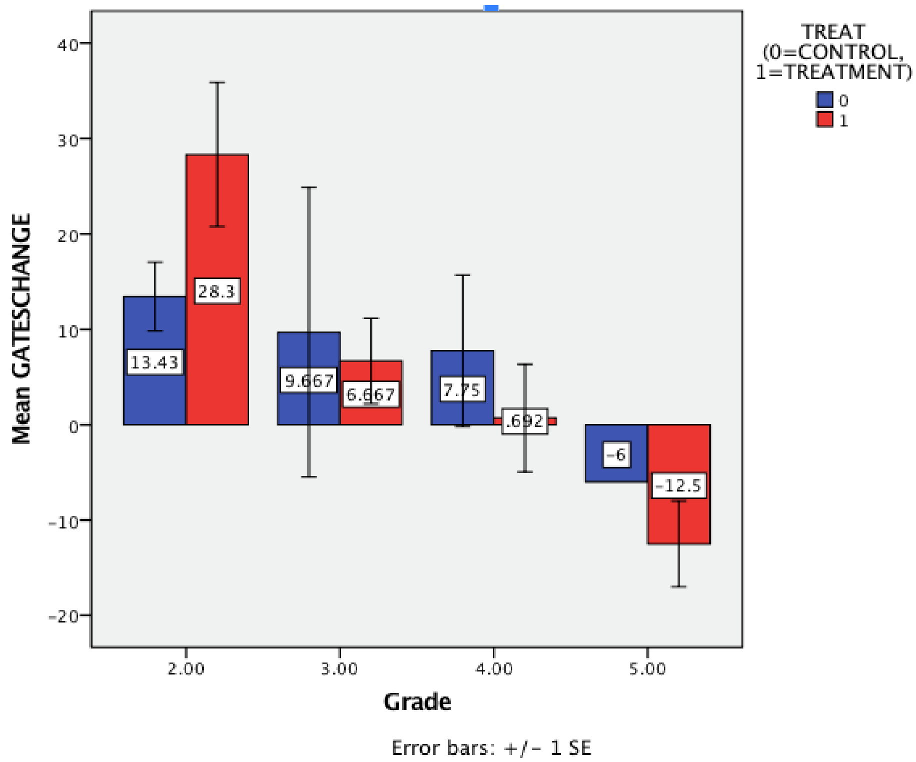

Figure 5. Gates Change Score by Grade and Treatment Group.

Research Question 2: What child-level factor or factors best predict benefit from the intervention?

To answer this question, we return to the model in which the outcome measure was accuracy on questions answered during the intervention. The decision was made to answer this question using the intervention outcome measure, as opposed to the Transfer stage accuracy or the Gates change score, because the largest intervention effect was observed during the intervention. In Model 1, the QRI decoding by treatment interaction was the only significant interaction at the child level. Again, this suggested that children who were poor decoders benefitted more from the intervention than better decoders (See Figure 1). A second model was run that included Gates baseline reading comprehension ability in the model. What entering this variable produced was a model where the main effect of QRI was no longer significant (p=.198) and the Treatment by QRI interaction was no longer significant (p=.827). However, the main effect of the Gates baseline score and the Treatment by Gates baseline score were both significant (ps<.025). This is perhaps not surprising, as the Gates score shares variance with the decoding variable (r=.58, p<.001), as predicted by the SVR model. The interaction effect can be seen in Figure 6.

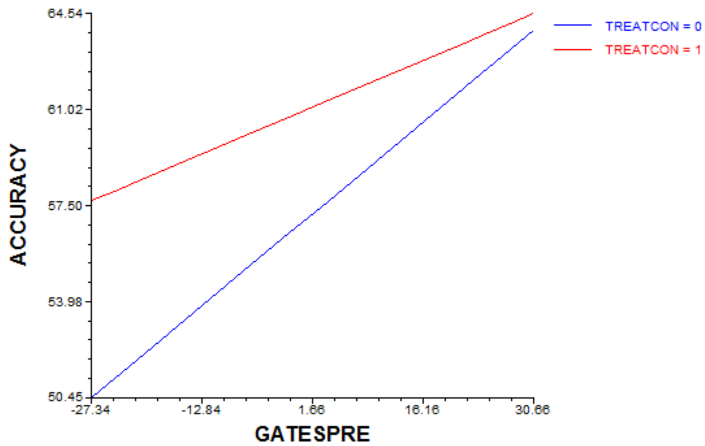

Figure 6. Treatment by Gates Baseline Reading Interaction.
Note. 0=control (blue); 1=intervention (red)

If we look to the Gates Change model for additional information, we can also conclude that the treatment tends to work best in producing longer-term transfer effects in younger children (i.e. 2nd graders), but is not effective among older children in producing this effect (See Figure 5).

Results – Study 2

Research Question 1) What are the relationships among motor, language, cognitive, and reading variables in DLL children?

Descriptive statistics were first run to determine the means and standard deviations of performance on the various motor tasks. The norms available for the DDK tasks from Fletcher (1972) were included in the table in order to compare performance. DDK statistics refer to syllable durations in milliseconds. Tapping statistics refer to inter-tap intervals in milliseconds. See Table 10 for means and standard deviations for all motor tasks.

Table 10

Oral and manual fine motor performance among 2nd-5th grade DLLs

Age	Task	Mean	Standard Deviation	Fletcher (1972) Mean	Fletcher (1972) St. Dev.
7 year old (*n*=7)	DDK /pa/	194.79	21.03	240.00	50.00
	DDK /pata/	192.49	32.04	253.33	86.67
	Repetitive Tapping	219.32	28.37		
	Alternating Tapping	426.15	144.74		
8 year old (*n*=18)	DDK /pa/	203.66	12.57	210.00	35.00
	DDK /pata/	209.94	45.07	206.67	60.00
	Repetitive Tapping	203.59	22.39		
	Alternating Tapping	361.92	114.40		
9 year old (*n*=21)	DDK /pa/	185.31	32.35	200.00	30.00
	DDK /pata/	187.17	40.61	196.67	53.33
	Repetitive Tapping	182.83	16.79		
	Alternating Tapping	308.13	96.41		
10 year old (*n*=8)	DDK /pa/	176.74	18.63	185.00	20.00
	DDK /pata/	167.12	41.05	183.33	50.00
	Repetitive Tapping	179.28	15.53		
	Alternating Tapping	268.04	76.46		

Note. All results reported in milliseconds. Two 11-year-old children were not included in these summary statistics. The *n* at each age level from the Fletcher (1972) study was 48.

In general, performance in each task tended to become quicker as children got older. There were no differences in performance between the repetitive (/pa/) vs. alternating (/pata/) oral motor task at any age. In fact, the means across the oral motor tasks and the repetitive tapping were near identical in their raw form. There tended to be higher variation among children in the alternating tasks than in the repetitive tasks.

Overall, children were significantly slower at performing the alternating tapping task than the repetitive tapping task ($t(55)=-10.12$, $p<.001$), which is expected in this age range.

Since many of the skills being examined in this data set improve across the developmental course, a table of partial correlations among all motor, language, cognitive, and reading variables controlling for the effect of age in months was generated. This resulted in Table 11 below.

Table 11

Partial Correlations and significance level controlling for age in months at start of intervention

	CELF-English	CELF-Spanish	QRI Decoding	Gates Baseline	Gates Change	English Artic Errors	Spanish Artic Errors	DDK /pa/	DDK /pata/	Repetitive Tapping	Alternating Tapping	WNV Cognitive
CELF-English	1											
CELF-Spanish	.382**	1										
QRI Decoding	.400**	.285*	1									
Gates Baseline	.469**	.366**	.583**	1								
Gates Change	-.186	-.340*	-.033	-.453**	1							
English Artic Errors	.011	.046	-.138	.125	.045	1						
Spanish Artic Errors	.041	-.508**	-.190	-.008	.136	.258	1					
DDK /pa/	-.108	-.342*	-.075	.064	.051	.082	.336*	1				
DDK /pata/	-.050	-.324*	-.122	-.172	.181	-.028	.196	.263	1			
Repetitive Tapping	-.154	-.058	-.098	.000	-.250	.128	.114	.222	.207	1		
Alternating Tapping	-.150	-.038	-.100	.050	.042	.295*	.088	-.027	.111	.390**	1	
WNV Cognitive	.274*	.006	-.022	.039	-.017	.143	.087	-.041	.452**	-.145	-.082	1

Note: * significant at $p<.05$; ** significant at $p<.01$

Significant and positive correlations were present among oral language (English and Spanish), decoding, and baseline reading ability. There was a strong, negative relationship between CELF-Spanish core language score and number of articulation errors in Spanish, but the parallel relationship did not exist in English. The average syllable duration of productions of /pa/ was negatively correlated with CELF-Spanish core language score (scores on CELF-Spanish increased as syllable duration of /pa/ decreased) but positively correlated with number of articulation errors in Spanish (faster productions of /pa/ correlated with more articulation errors in Spanish). The average syllable duration of productions of disyllable /pata/ was also negatively correlated with CELF-Spanish core language score (faster production of /pata/ related with higher CELF-Spanish scores). Neither of the oral fine motor tasks were associated with performance on CELF-English or articulation errors in English. The average syllable duration of /pata/ production was significantly and positively correlated with WNV scores which means that slower productions of /pata/ were related to higher WNV scores.

The repetitive tapping duration was significantly correlated with alternating tapping duration. Repetitive tapping duration was also correlated with the improvement on the Gates standardized reading test after the intervention (i.e. faster tapping speeds were associated with greater gain from the intervention), although the significance was borderline. The alternating tapping duration was significantly and positively correlated with the number of articulation errors in English (i.e. longer tapping intervals indicating slower alternating tapping speed associated with more errors in articulation). The oral motor and tapping tasks were not correlated with one another after age was controlled, although these correlations were significant before age was controlled.

Research Question 2) Do any motor variables predict performance on L1 or L2 language measures?

The first multiple linear regression equation was calculated to predict CELF-Spanish scores based on the four motor variables (/pa/, /pata/, repetitive tapping, and alternating tapping) and age in months. Each of these variables was grand-mean centered. A borderline significant regression equation was found ($F(5, 50) = 2.223$, $p = .066$) with an R^2 of .182. The resulting equation for predicting CELF-Spanish core language scores is as follows:

$$\hat{Y} = 79.7 - .18(Pa) - .10(Pata) + .06(RepetitiveTap) - .01(AlternatingTap) - .06(Age)$$

The unstandardized coefficients and significance levels are reported below.

Table 12

Coefficients for CELF-Spanish Prediction

	Beta (*B*)	Std. Error	t	Sig.
Constant	79.70	2.02	39.45	.000
/pa/	-0.18	0.08	-2.15	.036
/pata/	-0.10	0.05	-1.92	.061
Repetitive Tap	0.06	0.12	0.55	.585
Alternating Tap	-0.01	0.02	-0.35	.732
Age	-0.06	0.23	-0.26	.793

The constant or intercept is the mean CELF-Spanish score when all other variables in the equation are set to zero, which since they are centered, is also their mean. For each additional millisecond between a child's productions of /pa/, the CELF-Spanish score decreased by .18 points. This relationship was significant. The /pata/ predictor fell just short of reaching significance ($p = .061$), and indicated that for each additional millisecond between productions of syllables, the CELF-Spanish score decreased by .10 points. The tapping tasks were not significant predictors of CELF-Spanish score.

A second regression equation was modeled to evaluate if any motor variables predicted CELF-English score with age included in the model. Again all predictor variables were grand mean centered. The regression equation was not significant ($F(5, 50) = .43$, $p = .83$) with an R^2 of .04. The unstandardized coefficients and significance levels are reported below.

Table 13

Coefficients for CELF-English Prediction

	Beta (*B*)	Std. Error	t	Sig.
Constant	76.45	1.84	41.56	.000
Pa	-0.05	0.08	-0.63	.531
Pata	-0.00	0.05	0.04	.967
Repetitive Tap	-0.06	0.11	0.55	.585
Alternating Tap	-0.02	0.02	-0.78	.439
Age	-1.57	0.21	-0.76	.453

A final equation was generated to ensure that asymmetric language dominance or proficiency was not confounding the results. A new outcome variable, best language score, was calculated by choosing the higher of the two scores on the CELF-English and CELF-Spanish for each child. For 35 children, Spanish was the stronger language and, for the remaining 21, English was the stronger language. However, according to the parent questionnaire, all but four children heard Spanish at home. The average discrepancy between scores was 12.42 points (range 0-52). The regression equation was not significant ($F(5, 50) = 1.14$, $p = .35$) with an R^2 of .102. The unstandardized coefficients and significance levels are reported below.

Table 14

Coefficients for Best Language Prediction

	B	Std. Error	t	Sig.
Constant	84.29	1.68	50.16	.000
Pa	-0.14	0.07	-1.95	.056
Pata	-0.02	0.04	-0.56	.579
Repetitive Tap	0.0	0.10	0.02	.985
Alternating Tap	-0.01	0.02	-0.39	.699
Age	-0.19	0.19	-0.98	.334

Research Question 3) Do motor variables predict performance on *EMBRACE* comprehension questions?

Analytic strategy. Please see parallel section in Study 1 results.

Data treatment. Please see parallel section in Study 1 results.

Language and decoding predictors only model. The initial model was constructed such that only oral language (CELF-English and CELF-Spanish score) and decoding ability (QRI number of words decoded) were entered at the child level, in accordance with the SVR model of reading comprehension, which states that reading comprehension is a product of decoding and oral language comprehension (Hoover & Gough, 1990). This resulted in the following model, which I will refer to as the SVR model:

Level-1 Model

$ACCURACY_{ijk} = \pi_{0jk} + \pi_{1jk}*(PMIMCENT_{ijk}) + \pi_{2jk}*(TEXTTYPE_{ijk}) + \pi_{3jk}*(DIFFICUL_{ijk}) + e_{ijk}$

Level-2 Model

$\pi_{0jk} = \beta_{00k} + \beta_{01k}*(CELFENG_{jk}) + \beta_{02k}*(CELFSPA_{jk}) + \beta_{03k}*(QRI_{jk}) + r_{0jk}$
$\pi_{1jk} = \beta_{10k}$
$\pi_{2jk} = \beta_{20k}$
$\pi_{3jk} = \beta_{30k}$

Level-3 Model

$\beta_{00k} = \gamma_{000} + \gamma_{001}(GRADECEN_{k}) + u_{00k}$
$\beta_{01k} = \gamma_{010}$
$\beta_{02k} = \gamma_{020}$
$\beta_{03k} = \gamma_{030}$
$\beta_{10k} = \gamma_{100}$
$\beta_{20k} = \gamma_{200}$
$\beta_{30k} = \gamma_{300}$

ACCURACY$_{ijk}$ is the accuracy for child *j* in classroom *k* on the *i*th portion of the intervention (e.g. the PM stage of the easiest narrative text). Accuracy is a function of the coefficients of the three text-level factors (πs) plus an error term (e_{ijk}). π_{0jk} is the fitted mean for a given participant, which is also a function of the child-level factors plus a residual term (r_{0jk}). β_{00k} is the fitted mean for the classroom of students along with the classroom-level factor of grade and the error term (u_{00k}). γ_{000} is the fitted mean of the sample. γ_{001} is the main effect of grade. γ_{010} is the main effect of CELF-English. γ_{020} is the main effect of CELF-Spanish γ_{030} is the main effect of QRI decoding. γ_{100} is the main effect of intervention stage (PM or IM). γ_{200} is the main effect of text type (narrative or expository). γ_{300} is the main effect of text difficulty (Flesch-Kincaid grade level). The model allows for the partitioning of variance between text factors, child factors, and classroom factors. The results were as follows:

Table 15

Results of SVR model

Fixed Effect	Coefficient	Standard Error	*t*-ratio	Approx. *d.f.*	*p*-value
for INTERCPT 1, π_0					
for INTERCPT 2, β_{00}					
INTERCPT 3, γ_{000}	58.98	1.85	31.80	8	<.001
GRADE, γ_{001}	2.34	1.87	1.25	8	.256
for CELF-ENG, β_{01}					
INTERCPT 3, γ_{010}	0.26	0.11	2.37	43	.022
for CELFSPA, β_{02}					
INTERCPT 3, γ_{020}	-0.02	0.08	-0.24	43	.808
for QRI, β_{03}					
INTERCPT 3, γ_{030}	0.17	0.07	2.56	43	.014
for PMIM slope, π_1					
for INTERCPT 2, β_{10}					
INTERCPT 3, γ_{100}	-1.85	1.01	-1.83	598	.067
for TEXTTYPE slope, π_2					
for INTERCPT 2, β_{20}					
INTERCPT 3, γ_{200}	-23.03	1.13	-20.35	598	<.001
for DIFFICULTY slope, π_3					
for INTERCPT 2, β_{30}					
INTERCPT 3, γ_{300}	-4.38	0.55	-7.85	598	<.001

CELF-English, decoding ability, text type, and text difficulty were all significant predictors of performance during the intervention. This model resulted in a total deviance of 5413.46 with a total of 11 parameters estimated.

Subsequently, I entered the four motor predictors to the child level (i.e. Level 2) of the above equation. This resulted in the following fixed effects:

Table 16

Results of SVR + 4 Motor Predictors Model

Fixed Effect	Coefficient	Standard Error	*t*-ratio	Approx. *d.f.*	*p*-value
for INTERCPT 1, π_0					
for INTERCPT 2, β_{00}					
INTERCPT 3, γ_{000}	58.63	2.27	25.80	8	<.001
GRADE, γ_{001}	3.12	2.28	1.37	8	.208
for CELF-ENG, β_{01}					
INTERCPT 3, γ_{010}	0.23	0.10	2.38	39	.022
for CELFSPA, β_{02}					
INTERCPT 3, γ_{020}	0.07	0.08	0.90	39	.371
for QRI, β_{03}					
INTERCPT 3, γ_{030}	0.15	0.06	2.50	39	.017
for PA, β_{04}					
INTERCPT 3, γ_{040}	0.04	0.05	0.82	39	.420
for PATA, β_{05}					
INTERCPT 3, γ_{050}	0.07	0.03	2.49	39	.017
for REPETITIVE TAP, β_{06}					
INTERCPT 3, γ_{060}	-0.01	0.06	-0.11	39	.914
for ALTERNATE TAP, β_{07}					
INTERCPT 3, γ_{070}	0.00	0.01	0.12	39	.902
for PMIM slope, π_1					
for INTERCPT 2, β_{10}					
INTERCPT 3, γ_{100}	-1.85	1.01	-1.83	598	.067
for TEXTTYPE slope, π_2					
for INTERCPT 2, β_{20}					
INTERCPT 3, γ_{200}	-23.03	1.13	-20.35	598	<.001
for DIFFICULTY slope, π_3					
for INTERCPT 2, β_{30}					
INTERCPT 3, γ_{300}	-4.38	0.55	-7.85	598	<.001

The average syllable duration during productions of /pata/ was significant in predicting performance during the intervention. According to the coefficient, for each millisecond *increase* in intersyllable duration of /pata/, the accuracy improved by .07%. Overall, the model fit did not significantly improve over the SVR model (χ^2 (4) = 7.40, p = .115). However, when the non-significant motor predictors were removed from the equation for the sake of parsimony, leaving only /pata/, the significance of the /pata/ predictor increased (t(42)=2.80, p=.008) and the overall model fit was significantly better than the SVR model (χ^2 (1) = 6.77, p = .009).

Treatment condition (control=0, treatment=1) and a treatment by /pata/ interaction were subsequently added to the model at the child level to determine if success on the intervention depended in some way on motor skills. Treatment was borderline significant (t(40)=1.84, p=.07). The interaction of treatment by /pata/ was not significant (t(40)=.86, p=.39).

Research Question 4) Are there differences in motor performance between children who have language impairment and those with typical development?

In order to answer this question, children were qualified as presenting with language impairment (LI) based on a criterion of scoring at 75 or below on both the CELF-English and CELF-Spanish language test. If the child scored higher than 75 on either test, they were considered typically developing. Although 85 is the recommended cut score for determining the presence of LI (Semel, Wiig, & Secord, 2003; 2006), lower cut scores were used with this group of at-risk DLLs to avoid over-identification (Barragan et al., under review; Peña & Halle, 2011). This resulted in a total of 39 children with typical development and 17 children with language impairment.

A multivariate analysis of variance was conducted with each of the four motor tasks as dependent variables, language impairment status as the independent variable, and age in months as a covariate. There were no significant differences between groups on any of the four measures. See Table 17 for a summary of means, standard deviations, and *F*-tests.

Table 17

Summary of Motor Results Comparing Children with Typical Development and Language Impairment

	/pa/ Mean (SD)	/pata/ Mean (SD)	Repetitive tapping Mean (SD)	Alternating Tapping Mean (SD)
Typical Development	188.49 (27.92)	189.98 (41.06)	192.67 (25.43)	330.55 (107.73)
Language Impairment	200.06 (22.27)	198.50 (45.41)	193.23 (25.43)	336.01 (130.96)
F-test (1, 53)	2.45	.51	.03	.05
p-value	.12	.48	.86	.82

Note. Results are reported in milliseconds. For the oral motor tasks, this represents the average syllable duration. For the tapping tasks, this represents the average inter-tap duration.

Discussion – Study 1

The purpose of Study 1 was to evaluate the efficacy of the EMBRACE Spanish support intervention for at-risk dual language learners and to determine which individual characteristics of students make them more likely to benefit from the intervention. This goal was accomplished by comparing the control and intervention groups' performance on three different measures. The first of these was the performance on the intervention questions answered at the conclusion of each chapter, directly after either physically or imaginarily manipulating images on the iPad screen. This can be thought of as a proximal measure of intervention efficacy. The results suggested that, overall, EMBRACE was not effective over a control group at improving accuracy in all types of comprehension questions for multiple text types and difficulties. However, several interaction effects did point to potential benefit for some subgroups on some types of texts.

In terms of child-level variables, children who started off as lower decoders or lower comprehenders benefited more from the intervention than those that started off as more proficient in these skills. It is important to note that this sample was purposefully selected to include children with low reading comprehension skills, so the distribution is shifted to the lower range. Additionally, very poor decoders were not included in the study, so the lower end of the distribution of decoding skills in this sample could be described as having "adequate" decoding skills, although they still may have been considered below grade level. Therefore, the interaction of decoding by treatment effect suggests that the intervention was effective for a group of children with low to adequate decoding skills but less effective for average to good decoders. However, the baseline reading comprehension by treatment interaction suggests that the treatment was more effective for children with very low initial reading comprehension skills compared to those who started the intervention as better (but still low average) comprehenders.

Several built-in components of the intervention were meant to support poor decoders (i.e. preview of key vocabulary, clickable in-text words that could be heard in English along with Spanish translations, and highlighting of relevant objects in tandem with key words being read aloud). Perhaps this additional word-reading support facilitated comprehension in a subgroup of children whose primary barrier to successful reading comprehension was poor decoding skills. Proctor, Dalton, and Grisham (2007) found that struggling readers were more likely to access the digitally embedded features of a technology-based reading comprehension intervention than were stronger readers.

Considering this, it is also possible that weaker readers were more likely to access the built-in supports, and therefore receive more benefit, during the intervention.

In terms of text-level variables that interacted with treatment efficacy, for children in the intervention group, accuracy was higher for chapters read using the PM strategy than for the IM strategy. This finding is consistent with some previous studies with monolingual English speakers and Spanish-English dual language learners that have shown that accuracy tends to decrease in the IM stage (e.g. Adams, Restrepo, & Glenberg, under review; Glenberg, Goldberg, & Zhu, 2009; Walker, Adams, Restrepo, Fialko, & Glenberg, 2017). In addition, a larger treatment effect was seen for easier texts, but not for more difficult texts. The results suggest that new comprehension strategies should be taught embedded in texts at a difficulty level that represents the child's current reading level and not necessarily their current grade level. Gickling and Rosenfield (1995) suggest that accuracy should be 70-85% for practice reading exercises, and that if it is lower, the text difficulty is likely too high. An examination of the raw accuracy scores (See Appendix B) suggests that this accuracy standard was met, on average, for the three narrative texts, but not for the more difficult expository texts.

The second point of comparison between the intervention and control groups was the performance on the transfer texts. These were texts where all images were static and no prompting was given for strategy use, making these stories more similar to the kind of stories a child might read as part of their normal classroom curriculum. This can be thought of as a more distal measure of intervention efficacy. An interesting pattern of the effect of language ability emerged for this stage of the intervention. There was a strong, positive effect of CELF-English core language score on accuracy in this stage, with children earning the highest CELF-English scores outperforming children with the lowest CELF-English scores by nearly 30 percentage points. Conversely, there was a significant negative effect of CELF-Spanish such that children earning the highest CELF-Spanish scores had scores 20 percentage points *lower* than children earning the lowest CELF-Spanish scores (See Appendix C for a graphical representation of these trends). This pattern is likely indicative that children who are English-dominant tended to perform better on the transfer texts than children with weaker English language skills (who were likely more Spanish-dominant).

Several studies of the relationship between English and Spanish language variables and English reading comprehension suggest that English language variables are the strongest predictors of reading comprehension outcomes and often entirely moderate the relationship of Spanish language variables with English reading comprehension (Davison, Hammer, Lawrence, 2011; Gottardo & Mueller, 2009; Mancilla-Martinez & Lesaux, 2010; Manis, Lindsey, & Bailey, 2004). In the present study with this at-risk population with overall low language skills, it seems that English dominance was an especially important predictor of English reading comprehension accuracy. Although none of the child by treatment interactions were significant at this stage, the QRI decoding by treatment interaction was approaching significance, and considering the relatively small number of participants, may have not reached significance due to power issues. An examination of Figure 4 suggests a nearly 10-percentage point boost in accuracy for the lowest decoders as a result of the intervention. I would argue that this is an educationally relevant difference.

The final point of comparison between the control and intervention groups was on the change score in comprehension performance on the GMRT-4 standardized reading test. This can be considered the most distal measure of improvement in reading comprehension as a result of the intervention. There was no difference between the treatment and control groups overall for this measure. However, there was an average of a positive 15-point difference in Extended Scaled Scores for 2nd graders in the intervention group. This was likely due to the format of the Level 2 GMRT-4 test. At this level, as opposed to Levels 3-5, children are asked to choose a picture from a field of 3 that most closely corresponds to a sentence they have just read. This type of comprehension measure is much more closely aligned to the type of strategy taught during the intervention. Improvement among this second-grade group may be indicative of an improvement in the ability to create accurate mental imagery of written text, which is precisely what the IM stage of the intervention is designed to do.

For children in third through fifth grade, the GMRT-4 tests reading comprehension more traditionally by asking children to read a passage and answer written comprehension questions with no illustrations present. Figure 5 clearly demonstrates that, the higher the grade, the poorer the intervention group performed. In fact, the fifth graders demonstrated a negative change score from pre- to post-test. These results should be interpreted with caution, as there were only 3 children from 5th grade participating in this study (one in the control group and 2 in the intervention group). However, the trend of poorer performance with increasing grade was significant in the final model for this data set. Several factors that may have played a role include decreased motivation by the end of the intervention period and test fatigue as the post-test was also administered within one week of the state math and reading school-wide standardized tests. Again, Spanish CELF core language score was a borderline significant and negative predictor of Gates change score, suggesting that English dominance was likely to be associated with greater growth in reading comprehension.

The final research question related to which individual differences in students predicted success on the intervention. Across the intervention and transfer stages, we saw that lower decoding skills were related to greater treatment effects. When the Gates baseline reading ability was added to the model, the decoding variables were no longer significant. This is due to the fact that many of the same children who were poor decoders were also poor comprehenders, as one would expect based on the predictions of the SVR model. The interaction of treatment by baseline reading comprehension ability suggests that children who started the intervention as poor comprehenders were likely to benefit more from the intervention than those who started the intervention as better comprehenders.

The findings of this intervention study are compatible with previous research related to comprehension strategy interventions for DLLs with low language skills. Klingner and Vaughn (1996) used a reciprocal teaching intervention with DLLs in middle school with learning disabilities and found that students with adequate decoding but poor comprehension skills benefited the most from their 27-day intervention, although the overall treatment effect was not significant. Denton, Anthony, Parker, and Hasbrouck (2004) also found null treatment effects for improving comprehension for two reading tutoring programs (Read Well and Read Naturally) administered over a total of thirty 40-minute sessions over 10 weeks. These authors suggested that English vocabulary might have been a moderating factor that diminished treatment effects. Making measurable differences on complex constructs such as language proficiency and reading comprehension, especially among low-achieving, at-risk populations, is likely a longer-term endeavor than what can be captured in these short-term intervention studies.

Limitations

As is often the case in intervention studies, sample size was a limitation in this data set. Ideally, several hundred children would have participated to allow for more accurate estimates of the differences between control and treatment groups as well as the child by treatment interactions. The variability in this sample is also both a limitation and an asset. By recruiting such a variety of DLLs, including children diagnosed with language and reading disorders, those in SEI and dual language classrooms, in 2nd through 5th grade, with varying levels of ability and proficiency in each of their languages, it is difficult to make predictions about generalizability. However, this kind of variability is the reality of the DLL population, which I attempted to capture in the current project by measuring many of the factors that make this population so diverse. Future research that involves increased resources and access to a larger pool of participants may choose to focus on one sub-group (e.g. only those with diagnosed language impairment, only 2nd graders, or only Spanish-dominant students) in order to make more specific recommendations about the subgroups with which the intervention is effective.

Discussion – Study 2

The purpose of Study 2 was to determine the strength and direction of relationships between fine motor (oral and manual) skills, first and second language abilities, and reading abilities in DLLs in 2nd-5th grade. An examination of the descriptive statistics indicates that, in accordance with previous studies, older DLLs are faster at performing fine motor tasks than younger DLLs. Additionally, children's performance was more variable in alternating motor tasks than in repetitive tasks, which is also found in previous research (Denckla, 1973; Gasser, Rousson, Caflisch, & Jenni, 2010; Largo, et al., 2001). When comparing the averages and standard deviations of performance on the DDK tasks in the present sample to the norms available in Fletcher (1972), the performance looks quite similar, with a slight tendency for this group of DLLs to have shorter average syllable durations.

The correlations reported in Table 11 revealed that there were significant positive correlations among the oral language CELF scores in English and Spanish, the decoding variable, and baseline reading comprehension ability. These results are consistent with the SVR model (Hoover & Gough, 1990) of reading comprehension, and suggest that both oral language and decoding ability are related to reading comprehension. However, the strong and significant correlations between English and Spanish language ability (r =.382, p=.004) as well as between Spanish language ability and baseline reading comprehension (r=.366, p=.006) differ in strength and direction from some previous studies. Gottardo and Mueller (2009) found that correlations between English and Spanish vocabulary and syntactic abilities were positive, but not significant, and English oral language variables (but not Spanish) were significantly related to English reading comprehension. Proctor, August, Carlo, and Snow (2006) also found non-significant correlations between English reading comprehension and Spanish oral language variables, although they did find a small indirect effect of Spanish vocabulary predicting English reading comprehension using structural equation modeling.

Oral motor abilities, as measured by average syllable durations of productions of /pa/ and /pata/, were significantly related with Spanish (but not English) CELF core language scores. When using multiple regressions, similar results were found, in that oral motor ability (specifically average syllable durations of productions of /pa/) was a significant predictor of Spanish (but not English) oral language ability. Average syllable durations of /pa/ were also significantly and positively correlated with the number of articulation errors in Spanish, but not in English, which indicates that faster productions of /pa/ were related to fewer articulation errors in Spanish. When this information is taken together with the knowledge that, for all but 4 participants (see Appendix E, Table E1), Spanish is the native language, the pattern suggests that the relationship between motor and oral language abilities does exist for DLLs, but only as it relates to their first (or native) language. However, as addressed in research question 4 and summarized in Table 17, when an artificial binary group membership (typically developing or language impaired) was created in the data, there were no significant group differences on any of the measures. This is in contrast to previous studies that show a clear difference in fine motor performance for children with and without language impairment (Hill, 1998; Hill, 2001; Kooistra et al., 2005). It is important to note that dichotomizing a continuous variable is known to reduce power and reliability (MacCallum, Zhang, Preacher, & Rucker, 2002); therefore the relationship between oral motor abilities and Spanish language ability, namely that slower average syllable durations predict lower CELF-Spanish scores, can be considered more reliable. This relationship existed even though the range of scores on CELF-Spanish was relatively restricted to the lower end of the distribution (Range 40-108). It is possible, though, that this relationship was driven by low language proficiency and not an impairment of language ability.

Using Ullman and Pierpont's (2005) theory, if language and motor abilities are affected by deficiencies in the procedural memory system, and a second language is more likely to rely on declarative rather than procedural memory, this could explain the different relationships among motor and language abilities in this group of Spanish-native DLLs. Alternatively, the embodied interpretation of the BIA+ theory discussed in Adams (2016) would also predict stronger grounding of a native language, with a less proficient language being embodied to a lesser extent.

The relationships that did exist between motor and language abilities were only present for oral fine motor skills, but not manual fine motor skills. The average inter-tap durations on both the repetitive and alternating tapping tasks were not significantly related to any of the oral language measures. Brookman, McDonald, McDonald, and Bishop (2013) found that tapping speed was not affected by impairment in language or reading skills. However, in the present study, slower alternating tapping speeds were significantly correlated with more articulation errors in English. This could be partially due to the fact that during the scoring of the articulation errors, both accent-related errors and true articulation errors were counted as errors. As Nip and Blumenfeld (2015) report, speech motor patterns tend to be more variable in the second language than in the first, and less proficiency is associated with more articulation errors (Kormos & Dénes, 2004). According to Gasser and colleagues (2009), alternating tapping tasks are more difficult than repetitive tasks, and the most developmental progress is seen between ages 5-10. Therefore, the relationship between alternating tapping and English articulation errors may simply be indicative of the fact that alternating tapping and second language phonology are difficult, later-developing skills.

There was also a significant, positive relationship between cognitive abilities and average syllable durations of productions of /pata/. This correlation indicates that slower productions of /pata/ were related to higher cognitive scores. This relationship is puzzling because one would expect this correlation to be in the opposite direction. As Diamond (2000) suggests, fine motor and cognitive abilities develop in tandem and are likely products of the same cerebellar neural circuitry, so one would expect developmental progress in motor skills (i.e. faster productions) to be associated with higher cognitive scores.

When examining the relationship between motor abilities and performance on the *EMBRACE* intervention, addressing research question 3, average syllable duration of productions of /pata/ was the strongest predictor. Including /pata/ in the model to predict performance on the comprehension questions asked during the intervention improved model fit over a model that included only oral language and decoding variables, indicating that there may be some important relationship between oral motor and reading abilities. However, again the relationship was not in the expected direction. For each millisecond *increase* in average syllable duration of /pata/, the accuracy improved by .07%. It is possible that, since /pata/ and cognitive abilities were so highly correlated, the effect of /pata/ in the hierarchical linear model is actually just a proxy for the relationship between cognitive and reading abilities. This finding contrasts with Cutting and Scarborough (2006) who found that adding IQ to a model containing decoding and oral language did not improve model fit. However, Nation, Clarke, and Snowling (2002) reported a strong relationship between reading and cognitive abilities and demonstrated that most children who qualified as poor comprehenders (as many children in this sample were) also scored low on measures of general cognitive abilities. In contrast, this was not the case in this sample, as the correlation between cognitive and baseline reading ability was close to zero ($r = 0.039$).

The fact that baseline English reading and cognitive abilities were not correlated in this sample suggests that the more likely relationship present in these data is truly between motor and reading comprehension abilities. As Glenberg (2011) describes, *EMBRACE* was specifically designed to perpetuate use of the motor system while reading. By having children physically move (i.e. click and drag) objects on the screen or imagine such movement, the goal was to support simulations and the indexing of such movement to real-life experiences the child may have had. The speech motor articulatory systems are theorized to be a part of this indexing or grounding of language (Fadiga et al., 2002; Pulvermüller, 2006). Therefore, one would expect that better or more developed motor skills (i.e. faster productions of /pata/) would be related to better performance on comprehension in the intervention, while in reality the opposite effect was found. One possible explanation for the unexpected direction of the relationship between /pata/ and accuracy on comprehension questions during the intervention is that children who have weaker fine motor skills actually benefitted more from the intervention and were able to capitalize on the movement strategy support to answer more questions correctly. However, the non-significant interaction of /pata/ and treatment effect indicates that, even if children with weaker baseline oral fine motor skills benefitted more from the intervention, the effect was not significant in this sample. Nonetheless, the fact that oral fine motor performance predicts reading comprehension ability during an intervention designed to support use of the motor system suggests that the link between motor and reading abilities may be stronger than previously thought.

Limitations

Because this was an intensive intervention study, the number of children included was small for practical reasons. Certainly if the goal were to establish norms for motor performance in this population, one would plan to include a much larger number of children in order to make more reliable estimates. The diversity of language, cognitive, reading, and motor abilities in this sample was both an asset and a limitation. It allowed for capturing the variability that is inherently present in this population, but limited the ability to make conclusions about subgroups because of the small number of children.

Fine motor skill is by no means a unidimensional construct and the tasks chosen in the present study are just a small sample of possibilities for measuring fine motor skill. Additionally, the DDK and tapping tasks used are not, on their own, meaningful tasks for children. For example, Suggate and Stoeger (2014) advocate for excluding tapping tasks as a measure of fine motor control in favor of more meaningful manual tasks such as figure copying or block-building. Brookman and colleagues (2013) found different patterns of relationships between language, reading, and motor abilities depending on the extent to which the motor task required speed, sequencing or imitation and, as discussed earlier, did not find that tapping task speed was affected by language or reading impairment. As the Indexical Hypothesis postulates that meaning is grounded in real sensorimotor experiences (Glenberg & Robertson, 1999), the meaningfulness of the motor task may influence the extent to which language and motor performance are related. Future studies may choose to study a wider variety of motor tasks that require multiple fine motor skills and are more meaningful in order to specify the relationship between motor and language variables in DLLs.

Practical Implications

The strong relationship among both Spanish and English oral language, decoding, and English reading comprehension in this sample suggests that DLLs referred for concern about reading comprehension issues should be evaluated in both of their languages. In addition, an evaluation of fine motor skills may provide additional valuable information and may inform the type of interventions chosen to assist a child. In the case of dual language learners, practitioners may expect motor abilities to be more highly related to a child's native language than to their second language. Although, there were not group differences among children designated as typically developing or having language impairment in this sample, this difference may be present for more meaningful motor tasks. Interventions that incorporate movement and focus on connecting written language to real-world experience may be especially effective for children with low comprehension and low oral fine motor skills (Adams, 2016).

General Discussion

Efficacy of the Intervention

Overall, the *EMBRACE* intervention was marginally effective in improving comprehension for some types of texts and for some subgroups of children. While the main effect of the intervention did not reach the .05 level of significance in any stage of the intervention (PM, IM, transfer), an examination of the interaction terms revealed that children in the intervention group performed better on PM than on IM texts. Additionally, a treatment effect was seen for easier narrative and expository texts, but not for more difficult texts. There was an interesting interaction between decoding ability and treatment, indicating that children with lower decoding skills benefitted more from the intervention than children with better decoding skills, at least while reading stories using the PM or IM strategy.

There was little difference between the treatment and control groups on performance during the transfer stage suggesting that the children who did benefit from the intervention were not able to maintain effective strategy use in an unfamiliar story without prompting for physical or imagined movement. Similarly, there was no difference between control and intervention groups overall in improvement on a standardized reading comprehension measure, the GMRT-4. The most promising results came from the 2nd graders, who did show a 15-point advantage in the GMRT-4 extended scaled score as a result of having participated in the intervention. Children in 3rd-5th grade did not show any evidence of having benefitted from the intervention based on their change scores on the GMRT-4. However, making changes on standardized reading comprehension scores with relatively short comprehension strategy interventions is a notoriously difficult task for reading researchers (Droop, Elsacker, Voeten, & Verhoeven, 2016; Scammacca, Roberts, Vaughn, & Stuebing, 2015; Wassenburg, Bos, de Koning, & van der Schoot, 2015).

Language and Reading Predictors of Performance

Traditional models of reading comprehension such as the Simple View of Reading (Hoover & Gough, 1990; Gough & Tunmer, 1986) propose a strong role for both oral language and decoding skills in the prediction of reading comprehension. In the present study, child-level factors that were significant in predicting accuracy during the intervention included their English language ability and decoding skills. Children who had stronger skills in both areas were better comprehenders overall, when children in both the intervention and control group were considered. However, as mentioned earlier, children who started off with lower skills in decoding and comprehension benefitted more from the intervention. This finding is especially promising because it shows that even children with beginning decoding skills can benefit from comprehension strategy instruction.

In the transfer phase, which was considered a more distal measure of comprehension performance, a different pattern in the importance of language ability emerged. English language ability was again a positive predictor of comprehension performance. However, while Spanish language ability had almost no impact on performance during the intervention, it was a significant negative predictor of performance during the transfer stage. This pattern suggested that English language dominance among this group of DLLs with low overall language and/or reading skills was important for English reading comprehension performance when the intervention strategy support was removed. Similarly, Proctor, August, Carlo, and Snow (2005) found that English vocabulary knowledge (a proxy for proficiency) was crucial to English reading comprehension for native Spanish-speakers. There was also a negative relationship between Spanish language ability and amount of improvement as a result of the intervention, suggesting again that English oral language skills are key to successful growth in English reading comprehension.

Motor Predictors of Performance

Although motor skills have not been frequently included in studies of reading comprehension, there is ample evidence to suggest that a connection exists between these two sets of skills, as discussed previously. Additionally, the Indexical Hypothesis is based on the premise that motor skills are important for language processing. Therefore, a thorough examination of the relationship between motor skills and language and reading abilities was conducted in Study 2. Importantly, motor skills were only significantly related to Spanish (but not English) skills in this group of Spanish-dominant dual language learners. An effect of oral fine motor skill on reading comprehension accuracy was observed during the intervention. Adding the duration of syllables during productions of /pata/ explained additional variance in reading comprehension performance over and above a model that included oral language and decoding variables. These findings suggest a potentially important role for motor skills in reading comprehension, although larger studies are needed to establish the practical and educational relevance of this role.

References

Aaron, P. G., Joshi, M., & Williams, K. A. (1999). Not all reading disabilities are alike. *Journal of Learning Disabilities*, *32*(2), 120-137.

Adams, A. M. (2016). How language is embodied in bilinguals and children with specific language impairment. *Frontiers in psychology*, *7*, 1209.

Adams, A.M., Restrepo, M.A., & Glenberg, A.M. (under review). *Moved by Reading* in a DLL Spanish-speaking population. Manuscript submitted for publication.

Al Otaiba, S., Puranik, C., Zilkowksi, R., & Curran, T. (2009). Effectiveness of early phonological awareness interventions for students with speech or language impairments. *Journal of Special Education*, *43*, 107–128.

American Psychiatric Association (2000). Diagnostic and Statistical Manual of Mental Disorders-Text Revision, (4th ed.). Washington, DC: Author.

Amiel-Tison, C., Njiokiktjien, C., Vaivre-Douret, L., Verschoor, C. A., Chavanne, E., & Garel, M. (1996). Relation of early neuromotor and cranial signs with neuropsychological outcome at 4 years. *Brain and Development*, *18*(4), 280-286.

Arizona Department of Education (2017). Percentage of children approved for free or reduced-price lunches for school year 2017. Retrieved from https://cms.azed.gov/home/GetDocumentFile?id=5890e0bfaadebe05c0974e55. June 15, 2017.

August, D., & Shanahan, T. (2006). Developing Literacy in Second-Language Learners: A Report of the National Literacy Panel on Language Minority Children and Youth. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Barragan, B., Castilla-Earls, A., Martinez-Nieto, L., Restrepo, M.A., & Gray, S. (2017). *Performance of low-income dual language learners attending English-only schools on the CELF-4 Spanish.* Manuscript submitted for publication.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(04), 637-660.

Bergen, B., Lau, T. T. C., Narayan, S., Stojanovic, D., & Wheeler, K. (2010). Body part representations in verbal semantics. *Memory & cognition*, *38*(7), 969-981.

Berger, S. E. (2010). Locomotor expertise predicts infants' perseverative errors. *Developmental psychology*, *46*(2), 326.

Bergmann, K., & Macedonia, M. (2013, August). A virtual agent as vocabulary trainer:

Iconic gestures help to improve learners' memory performance. In *International Workshop on Intelligent Virtual Agents* (pp. 139-148). Springer Berlin Heidelberg.

Bishop, D. V. (1994). Is specific language impairment a valid diagnostic category? Genetic and psycholinguistic evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *346*(1315), 105-111.

Bishop, D. V. (2002). Motor immaturity and specific speech and language impairment: Evidence for a common genetic basis. *American Journal of Medical Genetics Part A*, *114*(1), 56-63.

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer* [Computer program], Version 6.0.26.

Brookman, A., McDonald, S., McDonald, D., & Bishop, D. V. (2013). Fine motor deficits in reading disability and language impairment: same or different?. *PeerJ*, *1*, e217.

Calderón, M. E. (2007). Teaching reading to English language learners, Grades 6–12: A framework for improving achievement in the content areas. Thousand Oaks, CA: Corwin Press.

Cameron, C. E., Brock, L. L., Murrah, W. M., Bell, L. H., Worzalla, S. L., Grissmer, D., & Morrison, F. J. (2012). Fine motor skills and executive function both contribute to kindergarten achievement. *Child development*, *83*(4), 1229-1244.

Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of learning disabilities*, *36*(2), 151-164.

Cermak, S. A., Ward, E. A. and Ward, L. M. (1986). The relationship between articulation disorders and motor coordination in children. *American Journal of Occupational Therapy*, *40*, 546–550.

Cheung, A. C., & Slavin, R. E. (2012). Effective reading programs for Spanish-dominant English language learners (ELLs) in the elementary grades: A synthesis of research. *Review of Educational Research*, *82*(4), 351-395.

Connor, C. M., Phillips, B. M., Kaschak, M., Apel, K., Kim, Y. S., Al Otaiba, S., ... & Lonigan, C. J. (2014). Comprehension tools for teachers: Reading for understanding from prekindergarten through fourth grade. *Educational Psychology Review*, *26*(3), 379-401.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*(2), 222-251.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific studies of reading*, *10*(3), 277-299.

Davison, M. D., Hammer, C., & Lawrence, F. R. (2011). Associations between preschool language and first grade reading outcomes in bilingual children. *Journal of Communication Disorders*, *44*(4), 444-458.

De Grauwe, S., Willems, R. M., Rueschemeyer, S. A., Lemhöfer, K., & Schriefers, H. (2014). Embodied language in first-and second-language speakers: Neural correlates of processing motor verbs. *Neuropsychologia*, *56*, 334-349.

Denckla, M. B. (1973). Development of speed in repetitive and successive finger-movements in normal children. *Developmental Medicine & Child Neurology*, *15*(5), 635-645.

Denton, C. A., Anthony, J. L., Parker, R., & Hasbrouck, J. E. (2004). Effects of two tutoring programs on the English reading development of Spanish-English bilingual students. *The Elementary School Journal*, *104*(4), 289-305.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, *5*(03), 175-197.

Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*, *54*(4), 1077-1088.

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second- language learners. *Reading Research Quarterly*, *38*(1), 78-103.

Durgunoğlu, A. Y. (2002). Cross-linguistic transfer in literacy development and implications for language learners. *Annals of Dyslexia*, *52*(1), 189-204.

Fernandes, B., Kester, E., Bauman, M. & Prath, S. (2011). Bilingual Articulation Phonology Assessment. [Mobile Application Software]. Retrieved from: https://itunes.apple.com/us/app/bilingual-articulation-phonology-assessment/id460830225?mt=8.

Fletcher, S.G. (1972) Time-by-count measurement of diadochokinetic rate. *Journal of Speech and Hearing Research*, 15, 763-770.

Gallese, V. (2007). Before and below theory of mind: Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society of*

London B, Biological Sciences, 362(1480): 659-669.

Gallese V. (2008). Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social Neuroscience, 3,* 317-333.

Gallese, V. & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in reason and language. *Cognitive Neuropsychology, 22*(3): 455-479.

Gasser, T., Rousson, V., Caflisch, J., & Jenni, O. G. (2010). Development of motor speed and associated movements from 5 to 18 years. *Developmental medicine & child neurology, 52*(3), 256-263.

Genesee, F., Geva, E., Dressler, C., & Kamil., M. L. (2006) Synthesis: Cross-linguistic relationship. In August, D. and Shanahan, T. (eds). *Developing literacy in second-language learners: Report of the national literacy panel on language minority children and youth*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Gickling, E., & Rosenfield, S. (1995). Best practices in curriculum-based assessment. *Best practices in school psychology*, *3,* 587-596.

Glenberg, A. M. (2011). How reading comprehension is embodied and why that matters. *International Electronic Journal of Elementary Education*, *4*(1), 5.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558-565.

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse processes*, *28*(1), 1-26.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*(3), 379-401.

Glenberg, A., Willford, J., Gibson, B., Goldberg, A., & Zhu, X. (2012). Improving reading to improve math. *Scientific Studies of Reading*, *16*(4), 316-340.

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading cinnamon activates olfactory brain regions. *Neuroimage*, *32*(2), 906-912.

Gottardo, A., & Mueller, J. (2009). Are first-and second-language factors related in predicting second-language reading comprehension? A study of Spanish-speaking children acquiring English as a second language from first to second grade. *Journal of Educational Psychology*, *101*(2), 330.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6-10.

Gualtieri, C. T., & Johnson, L. G. (2006). Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Archives of Clinical Neuropsychology*, *21*(7), 623-643.

Haager, D., & Windmueller, M. (2001). Early literacy intervention for English language learners at-risk for learning disabilities: Student and teacher outcomes in an urban school. *Learning Disability Quarterly*, *24*(4), 235-250.

Harris, C. L., Aycicegi, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics*, *24*(04), 561-579.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301-307.

Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., & Davidson, R. J. (2010). Cosmetic use of botulinum toxin-A affects processing of emotional language. *Psychological Science*, *21*(7), 895-900.

Henderson, S. E., & Barnett, A. L. (1992). *Movement assessment battery for children*. London: Psychological Corporation.

Hernandez, D. J., & Napierala, J. S. (2014). Early education, poverty, and parental circumstances among Hispanic children: pointing toward needed public policies. *Association of Mexican American Educators Journal*, *7*(2).

Hill, E. L. (1998). A dyspraxic deficit in specific language impairment and developmental coordination disorder? Evidence from hand and arm movements. *Developmental Medicine and Child Neurology*, *40*, 388–395.

Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders*, *36*(2), 149-171.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127-160.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of child language*, *37*(02), 229-261.

Jared, D., Poh, R. P. Y., & Paivio, A. (2013). L1 and L2 picture naming in Mandarin–English bilinguals: A test of bilingual dual coding theory. *Bilingualism: Language* *and Cognition*, *16*(02), 383-396.

Jiménez, R. T. (1997). The Strategic Reading Abilities and Potential of Five Low-Literacy Latina/o Readers in Middle School. *Reading Research Quarterly*, *32*(3), 224-243.

Jiménez, R. T., García, G. E., & Pearson, P. D. (1996). The reading strategies of bilingual Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly*, *31*(1), 90-112.

Kail, R. (1994). A method for studying the generalized slowing hypothesis in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *37*(2), 418-421.

Kirby, A., & Sugden, D. A. (2007). Children with developmental coordination disorders. *Journal of the Royal Society of Medicine*, *100*(4), 182-186.

Klingner, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *The Elementary School Journal*, *96*(3), 275-293.

Klingner, J. K., Artiles, A. J., & Barletta, L. M. (2006). English language learners who struggle with reading: Language acquisition or LD?. *Journal of Learning Disabilities*, *39*(2), 108-128.

Kochhar, R., & Fry, R. (2014). Wealth inequality has widened along racial, ethnic lines since end of Great Recession. *Pew Research Center*, *12*.

Kooistra, L., Crawford, S., Dewey, D., Cantell, M., & Kaplan, B. J. (2005). Motor correlates of ADHD: Contribution of reading disability and oppositional defiant disorder. *Journal of Learning Disabilities*, *38*(3), 195-206.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145-164.

Lakoff, G. (1987). Women, fire and dangerous things: What categories tell us about the nature of thought. Chicago: University of Chicago Press.

Largo, R. H., Caflisch, J. A., Hug, F., Muggli, K., Molnar, A. A., Molinari, L., ... & Gasser, T. (2001). Neuromotor development from 5 to 18 years. Part 1: timed performance. *Developmental Medicine & Child Neurology*, *43*(7), 436-443.

Leaders Project (January 8, 2014) Test review: CELF-4 Spanish. Retrieved from http://www.leadersproject.org/2014/01/08/test-review-celf-4-spanish/.

Leaders Project (July 26, 2013). Test review: CELF-4. Retrieved from http://www.leadersproject.org/2013/07/26/test-review-celf-4/.

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Economic Policy Institute, 1660 L Street, NW, Suite 1200, Washington, DC 20036.

Leonard, L. B., McGregor, K. K., & Allen, G. D. (1992). Grammatical morphology and speech perception in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *35*(5), 1076-1085.

Leslie, L., & Caldwell, J. (2010). *Qualitative Reading Inventory, 5th edition.* Boston, MA: Pearson Education.

Linan-Thompson, S., Vaughn, S., Hickman-Davis, P., & Kouzekanani, K. (2003). Effectiveness of supplemental reading instruction for second-grade English language learners with reading difficulties. *The Elementary School Journal*, *103*(3), 221-238.

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology, 95*(3), 482.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological methods*, *7*(1), 19.

Macedonia, M. (2014). Bringing back the body into the mind: gestures enhance word learning in foreign language. *Frontiers in Psychology, 5,* 1467.

Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education*, *5*(4), 196-211.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie reading tests* (4th ed.). Itasca, IL: The Riverside Publishing Company.

Mancilla-Martinez, J., & Lesaux, N. K. (2010). Predictors of reading comprehension for struggling readers: The case of Spanish-speaking language minority learners. *Journal of Educational Psychology*, *102*(3), 701.

Manis, F. R., Lindsey, K. A., & Bailey, C. E. (2004). Development of Reading in Grades K–2 in Spanish- Speaking English- Language Learners. *Learning Disabilities Research & Practice*, *19*(4), 214-224.

Maouene, J., Hidaka, S., & Smith, L. B. (2008). Body Parts and Early- Learned Verbs. *Cognitive Science*, *32*(7), 1200-1216.

Marley, S. C., Levin, J. R., & Glenberg, A. M. (2007). Improving Native American children's listening comprehension through concrete representations. *Contemporary Educational Psychology*, *32*(3), 537-550.

Marr, D., Cermak, S., Cohn, E. S., & Henderson, A. (2003). Fine motor activities in Head Start and kindergarten classrooms. *American Journal of Occupational Therapy*, *57*(5), 550-557.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Current opinion in neurobiology*, *11*(2), 194-201.

Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current Biology*, *25*(4), 530-535.

Mayer, M. (1967). *A boy, a Dog, a Frog, and a Friend.* New York, NY: Dial Press.

Mayer, M. (1967). *Frog on His Own*. New York, NY: Dial Press.

McPhillips, M., & Jordan- Black, J. A. (2007). The effect of social disadvantage on motor development in young children: a comparative study. *Journal of Child Psychology and Psychiatry*, *48*(12), 1214-1222.

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.

Moyer, A. (2004). Age, accent and experience in second language acquisition. Clevedon: Multilingual Matters.

Nag- Arulmani, S., Reddy, V., & Buckley, S. (2003). Targeting phonological representations can help in the early stages of reading in a non- dominant language. *Journal of Research in Reading*, *26*(1), 49-68.

Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2007). A longitudinal analysis of English language learners' word decoding and reading comprehension. *Reading and Writing*, *20*(7), 691-719.

Nation, K., Clarke, P., & Snowling, M. J. (2002). General cognitive ability in children with reading comprehension difficulties. *British Journal of Educational Psychology*, *72*(4), 549-560.

Neuman, S. B. (2008). *Educating the Other America*. Baltimore, MD: Brookes.

Nip, I. S., & Blumenfeld, H. K. (2015). Proficiency and linguistic complexity influence speech motor control and performance in Spanish language learners. *Journal of Speech, Language, and Hearing Research*, *58*(3), 653-668.

Norbury, C. F., Bishop, D. V., & Briscoe, J. (2001). Production of English finite verb morphology: A comparison of SLI and mild-moderate hearing impairment. *Journal of Speech, Language, and Hearing Research*, *44*(1), 165-178.

Peña, E. D., & Halle, T. G. (2011). Assessing preschool dual language learners: Traveling a multiforked road. *Child Development Perspectives*, *5*(1), 28-32.

Perie, M., Moran, R., & Lutkus, A. D. (2005). *NAEP 2004 trends in academic progress: Three decades of student performance in reading and mathematics*. National Center for Education Statistics, US Department of Education, Institute of Education Sciences.

Peter, B., Button, L., Stoel-Gammon, C., Chapman, K., & Raskind, W. H. (2013). Deficits in sequential processing manifest in motor and linguistic tasks in a multigenerational family with childhood apraxia of speech. *Clinical linguistics & phonetics*, *27*(3), 163-191.

Peter, B., Matsushita, M., & Raskind, W. H. (2012). Motor sequencing deficit as an endophenotype of speech sound disorder: a genome-wide linkage analysis in a multigenerational family. *Psychiatric genetics*, *22*(5), 226.

Peter, B., Wijsman, E. M., Nato Jr, A. Q., Matsushita, M. M., Chapman, K. L., Stanaway, I. B., ... & University of Washington Center for Mendelian Genomics. (2016). Genetic candidate variants in two multigenerational families with childhood apraxia of speech. *PloS one*, *11*(4), e0153864.

Planty, M., Hussar, W., Snyder, T., Provasnik, S., Kena, G., Dinkes, R., et al. (2008). The Condition of Education (NCES 2008-031). Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Depart- ment of Education.

Porter, A. (2016). A helping hand with language learning: teaching French vocabulary with gesture. *The Language Learning Journal*, *44*(2), 236-256.

Preissl, H., Pulvermüller, F., Lutzenberger, W., & Birbaumer, N. (1995). Evoked potentials distinguish between nouns and verbs. *Neuroscience Letters*, *197*(1), 81-83.

Proctor, C. P., August, D., Carlo, M. S., & Snow, C. (2006). The intriguing role of Spanish language vocabulary knowledge in predicting English reading comprehension. *Journal of Educational Psychology*, *98*(1), 159.

Proctor, C. P., Dalton, B., & Grisham, D. L. (2007). Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research, 39*(1), 71-93.

Pulvermüller, F., Härle, M., & Hummel, F. (2001). Walking or talking?: Behavioral and neurophysiological correlates of action verb processing. *Brain and Language, 78*(2), 143-168.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience, 21*(3), 793-797.

Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, 103*(20), 7865-7870.

Pulvermüller, F., Lutzenberger, W., & Preissl, H. (1999). Nouns and verbs in the intact brain: evidence from event-related potentials and high-frequency cortical responses. *Cerebral cortex, 9*(5), 497-506.

Raudenbush, S.W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.) Thousand Oaks, CA: Sage.

Restrepo, M. A., Morgan, G. P., & Thompson, M. S. (2013). The efficacy of a vocabulary intervention for dual-language learners with language impairment. *Journal of Speech, Language, and Hearing Research, 56*(2), 748-765.

Rintala, P., Pienimaki, K., Ahonen, T. & Cantell, M. (1998). Effects of a psychomotor training programme on motor skill development in children with developmental language disorders. *Human Movement Science, 17*, 721–737.

Robinson, R. J. (1991). Causes and associations of severe and persistent specific speech and language disorders in children. *Developmental Medicine and Child Neurology, 33*, 943–962.

Roth, M. A., McCaul, E., & Barnes, K. (1993). Who becomes an “at-risk” student? The predictive value of a kindergarten screening battery. *Exceptional Children, 59*(4), 348-358.

Semel, E., Wiig, E.H., Secord, W.A. (2006). *Clinical evaluation of language fundamentals* (4th ed.), Spanish Edition. San Antonio, TX: Harcourt Assessment, Inc.

Semel, E.,Wiig, E. H., & Secord, W. A. (2003*). Clinical evaluation of language* *fundamentals* (4th ed.). Toronto: The Psychological Corporation.

Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology*, *31*(5), 811.

Stillwell, R., & Sable, J. (2013). Public School Graduates and Dropouts from the Common Core of Data: School Year 2009-10. First Look (Provisional Data). NCES 2013-309rev. *National Center for Education Statistics*

Suggate, S. P., & Stoeger, H. (2014). Do nimble hands make for nimble lexicons? Fine motor skills predict knowledge of embodied vocabulary items. *First Language*, *34*(3), 244-261.

Swanson, H. L., Sáez, L., Gerber, M., & Leafstedt, J. (2004). Literacy and cognitive functioning in bilingual and nonbilingual children at or not at risk for reading disabilities. *Journal of Educational Psychology*, *96*(1), 3.

Tervo, R. C., Azuma, S., Fogas, B., & Fiechtner, H. (2002). Children with ADHD and motor dysfunction compared with children with ADHD only. *Developmental Medicine & Child Neurology*, *44*(6), 383-390.

Thordardottir, E., Cloutier, G., Ménard, S., Pelland-Blais, E., & Rvachew, S. (2015). Monolingual or bilingual intervention for primary language impairment? A randomized control trial. *Journal of Speech, Language, and Hearing Research*, *58*(2), 287-300.

Tymms, P. (2014). Effect sizes in multilevel models. In Schagen, I. and Elliot, K. (eds.). *But What Does it Mean?: The Use of Effect Sizes in Educational Research*. Upton Park, United Kingdom: National Foundation for Educational Reseearch.

Tymms, P., Merrell, C., & Henderson, B. (1997). The First Year at School: A Quantitative Investigation of the Attainment and Progress of Pupils. *Educational Research and Evaluation*, *3*(2), 101-118.

U.S. Department of Commerce, Census Bureau (2015). American Community Survey (ACS). *Digest of Education Statistics 2016*, table 102.60.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP, 2011 Reading Comprehension).

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP, 2015 Reading Comprehension).

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP, 2017).

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*(3), 399-433.

Ullman, M.T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory and practice* (pp. 141–178). Washington, DC: Georgetown University Press.

Van Gelderen, A., Schoonen, R., Stoel, R. D., De Glopper, K., & Hulstijn, J. (2007). Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology*, *99*(3), 477.

Vaughn, S., Mathes, P. G., Linan-Thompson, S., Cirino, P. T., Carlson, C. D., Pollard-Durodola, S. D., et al. (2006). First-grade English language learners at-risk for reading problems: Effectiveness of an English intervention. *Elementary School Journal*, *107*, 153–180.

Verhoeven, L. (2007). Early bilingualism, language transfer, and phonological awareness. *Applied Psycholinguistics*, *28*(03), 425-439.

Vukovic, N., & Shtyrov, Y. (2014). Cortical motor systems are involved in second-language comprehension: Evidence from rapid mu-rhythm desynchronisation. *NeuroImage*, *102*, 695-703.

Vukovic, N., & Williams, J. N. (2014). Automatic perceptual simulation of first language meanings during second language sentence processing in bilinguals. *Acta psychologica*, *145*, 98-103.

Walker, E., Adams, A., Restrepo, M. A., Fialko, S., & Glenberg, A. M. (2017). When (and how) interacting with technology-enhanced storybooks helps dual language learners. *Translational Issues in Psychological Science*, *3*(1), 66.

Wechsler, D., & Naglieri, J.A. (2006). *Wechsler nonverbal scale of ability*. San Antonio, TX: Harcourt Assessment.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature neuroscience*, *7*(7), 701-702.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock– Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological science*, *13*(2), 168-171.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1.

APPENDIX A

ENGLISH AND SPANISH LANGUAGE MODELS AND CORRELATION TABLES

Table A1

English language predictors only model

Fixed Effect	Coefficient	Standard Error	t-ratio	Approx. *d.f.*	*p*-value
CELF-English	0.268937	0.110589	2.432	42	0.019
MLU English	1.637465	1.197392	1.368	42	0.179
NDW English	-0.036885	0.078383	-0.471	42	0.640
Ungrammaticality English	-0.052251	0.113757	-0.459	42	0.648

Note. Coefficients refer to prediction of comprehension accuracy during the intervention stage.

Table A2

Correlation Matrix of English language predictors

	CELF-English	MLU English	NDW English	Ungrammaticality English
CELF-English	1			
MLU English	.298**	1		
NDW English	.444**	.511**	1	
Ungrammaticality English	-.313**	.094*	-.177**	1

Note. **Correlation is significant at the 0.01 level (two-tailed)
*Correlation is significant at the 0.05 level (two-tailed)

Table A3

Spanish Language Predictors Only Model

Fixed Effect	Coefficient	Standard Error	t-ratio	Approx. *d.f.*	*p*-value
CELF-Spanish	0.243111	0.132032	1.841	42	0.073
MLU Spanish	-1.515075	1.317217	-1.150	42	0.257
NDW Spanish	-0.003769	0.076980	-0.049	42	0.961
Ungrammaticality Spanish	0.117865	0.078324	1.505	42	0.140

Note. Coefficients refer to prediction of comprehension accuracy during the intervention stage.

Table A4

Correlation Matrix of Spanish Language Predictors

	CELF-Spanish	MLU Spanish	NDW Spanish	Ungrammaticality Spanish
CELF-Spanish	1			
MLU Spanish	.546**	1		
NDW Spanish	.516**	.616**	1	
Ungrammaticality Spanish	-.454**	.048	-.117	1

Note. **Correlation is significant at the 0.01 level (two-tailed)

APPENDIX B

RAW INTERVENTION PERFORMANCE

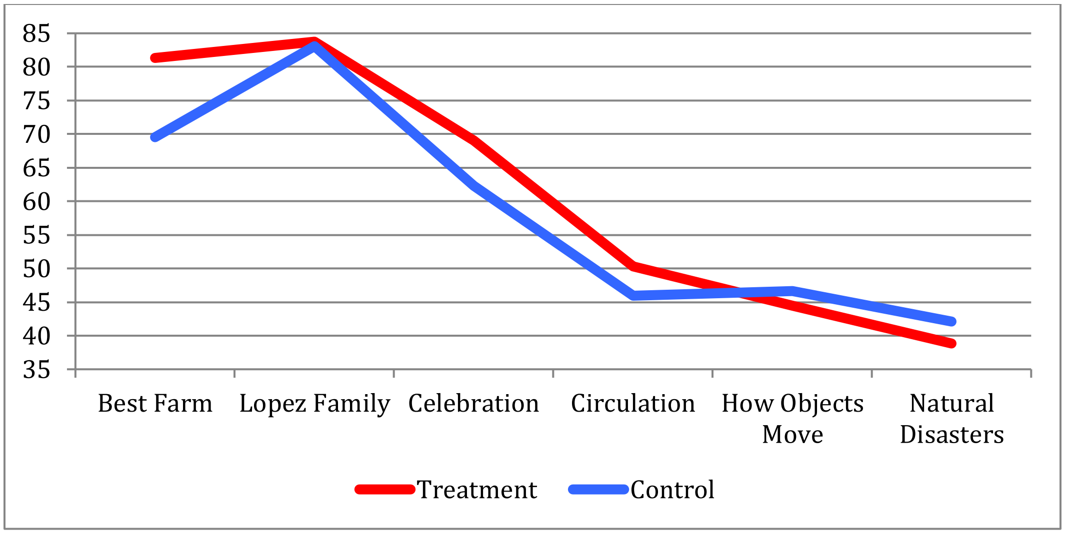

Figure B1. Raw PM Performance During Intervention.

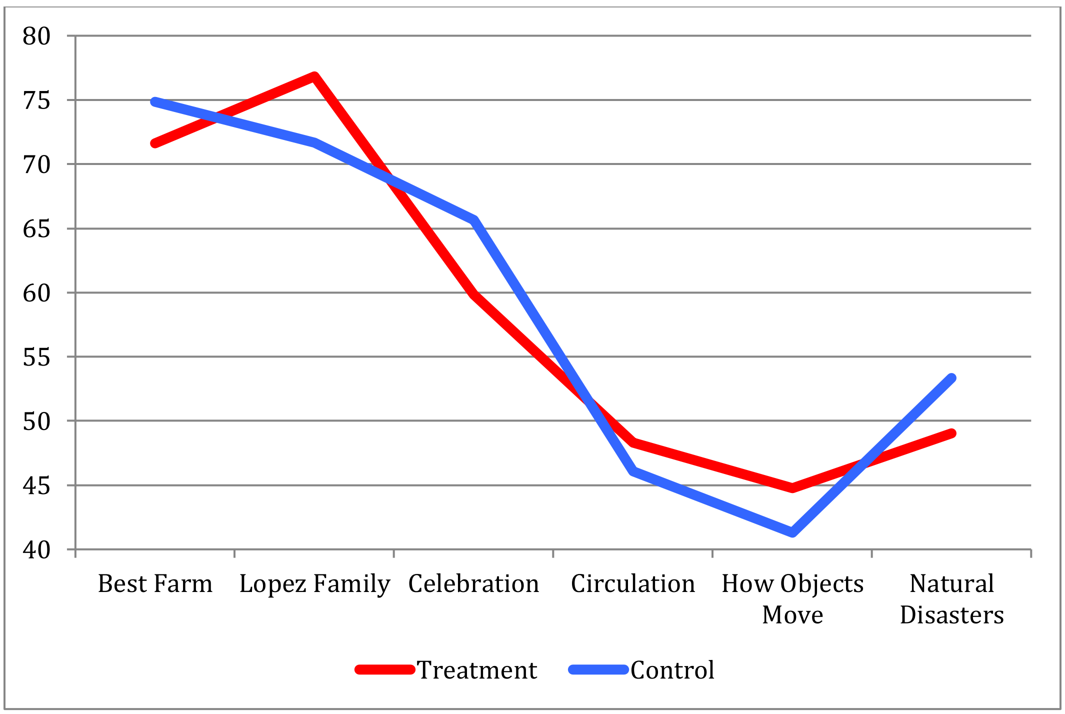

Figure B2. Raw IM Performance During Intervention.

APPENDIX C

EFFECT OF LANGUAGE ON TRANSFER ACCURACY

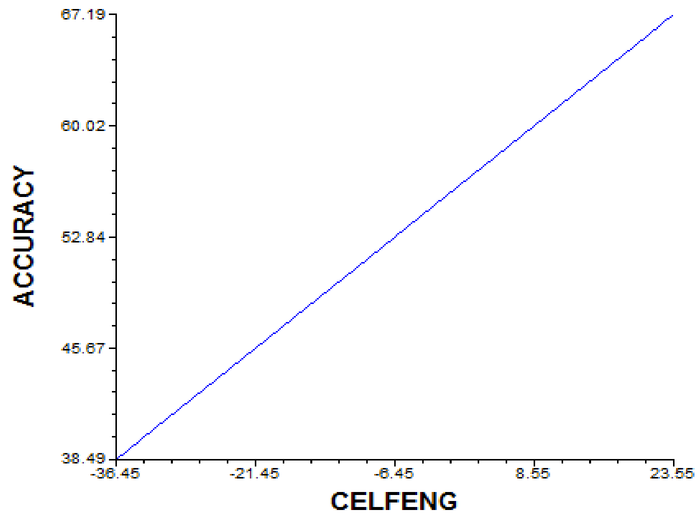

Figure C1. Effect of CELF-English on transfer stage comprehension question accuracy.

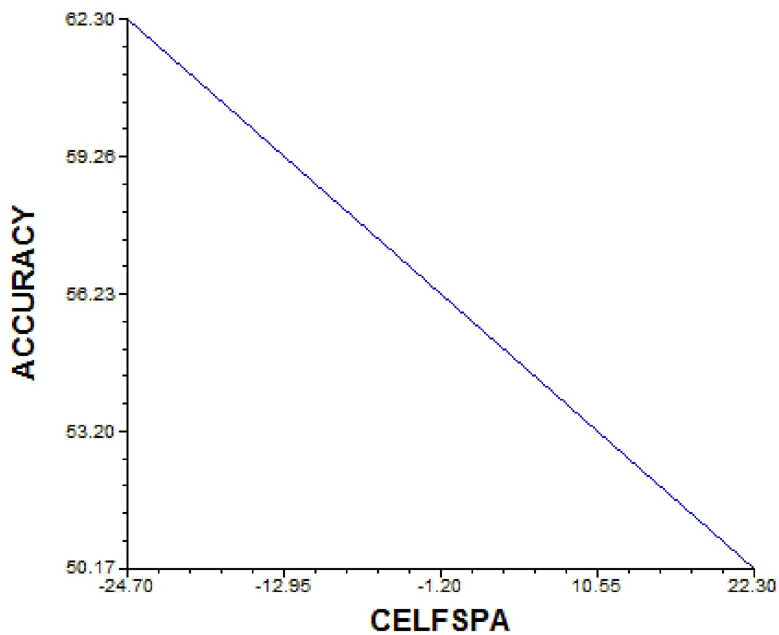

Figure C2. Effect of CELF-Spanish on transfer stage comprehension question accuracy

APPENDIX D

EFFECT SIZES SUMMARY

Table D1

Intervention Stage Dichotomous Variable Effect Sizes

Effect	Effect Size
Treatment	0.51
Intervention Stage (PM or IM)	0.04
Text Type (Narrative or Expository)	-1.78

Note. This effect size can be defined as the difference between the estimated means for the groups defined by the dummy codings 1 and 0 expressed as a fraction of the appropriate level standard deviation (level 1=text level, level 2=child level, level 3=classroom level), after controlling for covariates in the model (Tymms, et al., 1997, p.112). These effect sizes were calculated using the following formula: $\Delta = \beta_1/\sigma_e$ in accordance with Tymms, 2004 where β_1 is the coefficient and σ_e is the appropriate level standard deviation.

Table D2

Intervention Stage Continuous Variable Effect Sizes

Effect	Predictor Standard Deviation	Effect Size
Grade	1.16	2.13
CELF-English	13.4	1.06
CELF-Spanish	15.94	<.001
QRI Decoding	25.21	2.54
Text Difficulty	1.01	-0.49

Note. This effect size can be defined as the change in the outcome measure that will be produced by a change of +/- one standard deviation on the continuous predictor variable, standardized by the within classroom standard deviation adjusted for covariates in the model (Tymms, et al., 1997). These effect sizes were calculated using the following formula: $\Delta = 2*\beta_1*SD_{predictor}/\sigma_e$ in accordance with Tymms, 2004 where β_1 is the coefficient and σ_e is the appropriate level (level 1=text level, level 2=child level, level 3=classroom level) standard deviation.

Table D3

Intervention Stage Interaction Effect Sizes

Effect	Effect Size
Treatment by QRI @ 25th percentile	0.33
Treatment by QRI @ 75th percentile	-0.06
Treatment by Difficulty @ 25th Percentile	0.28
Treatment by Difficulty @ 75th Percentile	0.05
Treatment by Intervention Stage: PM	0.24
Treatment by Intervention Stage: IM	0.08

Note. Only significant or borderline significant interaction effects were considered. These effect sizes are expressed in Cohen's *d* format. In each case the effect size represents the difference in comprehension performance accuracy (in standard deviations) between the intervention and control group. Positive numbers indicate an advantage for the intervention group, while negative numbers indicate an advantage for the control group.

Table D4

Transfer Stage Dichotomous Variable Effect Sizes

Effect	Effect Size
Treatment	0.49
Text Type (Narrative or Expository)	-2.27

Note. This effect size can be defined as the difference between the estimated means for the groups defined by the dummy codings 1 and 0 expressed as a fraction of the appropriate level standard deviation (level 1, 2 or 3), after controlling for covariates in the model (Tymms, et al., 1997, p.112). These effect sizes were calculated using the following formula: $\Delta = \beta_1/\sigma_e$ in accordance with Tymms, 2004 where β_1 is the coefficient and σ_e is the appropriate level (level 1=text level, level 2=child level, level 3=classroom level) standard deviation.

Table D5

Transfer Stage Continuous Variable Effect Sizes

Effect	Predictor Standard Deviation	Effect Size
Grade	1.16	2.22
CELF-English	13.40	1.58
CELF-Spanish	15.94	-0.85
QRI Decoding	25.21	1.64

Note. This effect size can be defined as the change in the outcome measure that will be produced by a change of +/- one standard deviation on the continuous predictor variable, standardized by the within classroom standard deviation adjusted for covariates in the model (Tymms, et al., 1997). These effect sizes were calculated using the following formula: $\Delta = 2 * \beta_1 * SD_{predictor} / \sigma_e$ in accordance with Tymms, 2004 where β_1 is the coefficient and σ_e is the appropriate level (level 1=text level, level 2=child level, level 3=classroom level) standard deviation.

Table D6

Transfer Stage Interaction Effect Sizes

Effect	Effect Size
Treatment by QRI @ 25th percentile	0.39
Treatment by QRI @ 75th percentile	-0.05

Note. Only significant or borderline significant interactions were considered. These effect sizes are expressed in Cohen's *d* format. In each case the effect size represents the difference in comprehension performance accuracy (in standard deviations) between the intervention and control group. Positive numbers indicate an advantage for the intervention group, while negative numbers indicate an advantage for the control group.

Table D7

Gates Change Dichotomous Variable Effect Size

Effect	Effect Size
Treatment	0.05

Note. This effect size can be defined as the difference between the estimated means for the groups defined by the dummy codings 1 and 0 expressed as a fraction of the appropriate level standard deviation (level 1=child level, level 2=classroom level), after controlling for covariates in the model (Tymms, et al., 1997, p.112). These effect sizes were calculated using the following formula: $\Delta = \beta_1/\sigma_e$ in accordance with Tymms, 2004 where β_1 is the coefficient and σ_e is the appropriate level standard deviation.

Table D8

Gates Change Continuous Variable Effect Sizes

Effect	Predictor Standard Deviation	Effect Size
Grade	1.16	-3.05
CELF-English	13.40	1.31
CELF-Spanish	15.94	-1.35
QRI Decoding	25.21	0.91

Note. This effect size can be defined as the change in the outcome measure that will be produced by a change of +/- one standard deviation on the continuous predictor variable, standardized by the within classroom standard deviation adjusted for covariates in the model (Tymms, et al., 1997). These effect sizes were calculated using the following formula: $\Delta = 2*\beta_1*SD_{predictor}/\sigma_e$ in accordance with Tymms, 2004 where β_1 is the coefficient and σ_e is the appropriate level (level 1=child level, level 2=classroom level) standard deviation.

APPENDIX E

SELECTED PARENT QUESTIONNAIRE RESULTS

Table E1

Parent Questionnaire Language Proficiency Results

Question	Scale	Frequency
Are you worried about how your child speaks?	1=No they speak well, 2=Yes, a little worried, 3=Yes, I worry, 4=Yes, very worried	1=32 2=18 3=2 4=1 Missing=3
Does child say sentences that aren't grammatical or don't sound right?	Yes or No	Yes=21 No=26 Missing=9
Is there a history of speech, language or reading disability?	Yes or No	Yes=17 No=33 Missing=6
What language(s) are spoken in your home?	Fill in the Blank	Spanish only=34 English only=4 Both Spanish and English=13 Missing=5
How well does your child Speak English?	1=does not speak at all, 2=says some words and phrases, 3=can have a simple conversation, 4=speaks fluently with some errors, 5=speaks fluently	1=0 2=3 3=7 4=17 5=23 Missing=6
How well does your child Speak Spanish?	1=does not speak at all, 2=says some words and phrases, 3=can have a simple conversation, 4=speaks fluently with some errors, 5=speaks fluently	1=0 2=5 3=9 4=22 5=13 Missing=7
How much Spanish does the child HEAR at home during the week? FROM MOM	1=never, 2=sometimes, 3=most of the time, 4=all the time	1=2 2=4 3=13 4=35 Missing=2

How much English does the child HEAR at home during the week? FROM MOM	1=never, 2=sometimes, 3=most of the time, 4=all the time	1=12 2=20 3=11 4=10 Missing=3

Table E2

Parent Questionnaire Reading Results

Question	Scale	Frequency
In what languages does your child read?	Fill in the Blank	Spanish only=1 English only=34 Both Spanish and English=12 Missing=8
How well does your child read English?	1=well for his/her age, 2=with some difficulty, 3=he/she can not read well	1=20 2=27 3=3 Missing=6
How well does your child read Spanish?	1=well for his/her age, 2=with some difficulty, 3=he/she can not read well	1=7 2=23 3=18 Missing=8

Table E3

Parent Questionnaire SES Results

Question	Scale	Frequency
What is Mother's Highest Level of Education?	1=Elementary, 2=High School/GED 3=Some college/Technical School 4=Postgraduate/Professional	1=20 2=21 3=7 4=5 Missing=3
What is Father's Highest Level of Education?	1=Elementary, 2=High School/GED 3=Some college/Technical School 4=Postgraduate/Professional	1=20 2=20 3=4 4=1 Missing=11

APPENDIX F

SELECTED TEACHER QUESTIONNAIRE RESULTS

Table F1

Teacher Questionnaire Language and Literacy Proficiency Results

Question	Scale	Frequency
How often does this child speak to you in Spanish?	1=Never, 2=Occasionally, 3=Frequently, 4=Always	1=36 2=4 3=3 4=12 Missing=1
How often does this child speak to you in English?	1=Never, 2=Occasionally, 3=Frequently, 4=Always	1=0 2=3 3=3 4=49 Missing=1
How often does this child speak to classmates in Spanish outside of the classroom?	1=Never, 2=Occasionally, 3=Frequently, 4=Always	1=8 2=41 3=5 4=1 Missing=1
How often does this child speak to classmates in English outside of the classroom?	1=Never, 2=Occasionally, 3=Frequently, 4=Always	1=1 2=4 3=30 4=20 Missing=1
How would you rate this child's understanding of English?	1=Doesn't understand anything 2=Understands a little 3=Understands only the main ideas 4=Understands most of what is said 5=Understands like a native speaker	1=0 2=0 3=4 4=40 5=11 Missing=1
How would you rate this child's ability to speak English?	1=Cannot speak any English 2=Speaks a little English 3=Speaks limited English with errors 4=Speaks fluent English with errors 5=Speaks like a native speaker	1=0 2=1 3=4 4=45 5=5 Missing=1

Are you concerned about this child's oral language development?	1=Not concerned 2=Somewhat concerned 3=Very concerned	1=41 2=11 3=3 Missing=1
Are you concerned about this child's literacy development?	1=Not concerned 2=Somewhat concerned 3=Very concerned	1=11 2=30 3=14 Missing=1

APPENDIX G

EXAMPLE STORY

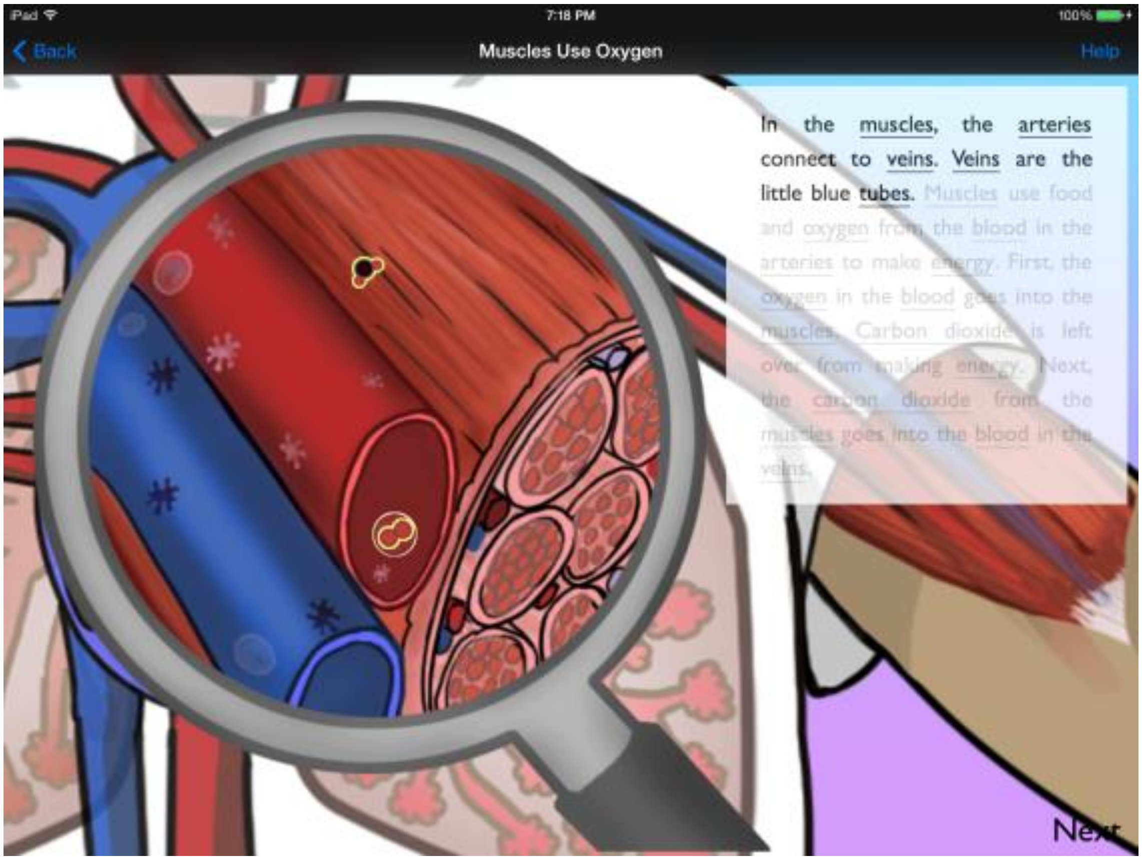

Figure G1. Example screenshot from the Circulation text.