Title
Examining the Accuracy of Teachers’ Ratings of Working Memory in English Learners: A Validation Study

Permalink
https://escholarship.org/uc/item/4wk5x44m

Author
Uppal, Harpreet Kaur

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Examining the Accuracy of Teachers' Ratings of Working Memory in English Learners: A Validation Study

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Education

by

Hapreet Kaur Uppal

June 2015

Dissertation Committee:
  Dr. H. Lee Swanson, Chairperson
  Dr. Marsha Ing
  Dr. Michael J. Orosco

The Dissertation of Harpreet Kaur Uppal is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. H. Lee Swanson, for his guidance, encouragement, and most of all patience throughout the entire process. I would also like to thank the rest of my dissertation committee: Dr. Marsha Ing and Dr. Michael J. Orosco for their encouragement and insightful comments.

I am indebted to Dr. Cathy Lussier for providing essential support and guidance during my graduate work. I am very grateful to Wenson, Mily, Jennifer, Sarana, Christy, Erik, and Andres for their friendship, encouragement, support, coffee sessions, and Tuesday lunches. A special thanks to my cohort mates, Truman, Allan, and Luyao for their words of encouragement and support throughout the graduate years. Also, I am thankful to Vanda Yamaguchi for assisting me with the administrative tasks necessary for completing my doctoral program.

I would like to thank my loving and caring family: my mom, Manjit Kaur Uppal, dad, Ujjagar Singh Uppal, siblings (Gurpreet, Sagandeep, and Pardeep), jeeju, cousins, aunts, uncles, dadi, and nani for always supporting me and encouraging me throughout this process and my life.

Above all, nothing would have been possible if it was not for Waheguru Ji's blessings and guidance.

DEDICATION

I dedicate my dissertation work to my parents for instilling the importance of hard work and higher education.

ABSTRACT OF THE DISSERTATION

Examining the Accuracy of Teachers' Ratings of Working Memory in English Learners: A Validation Study

by

Harpreet Kaur Uppal

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside, June 2015
Dr. H. Lee Swanson, Chairperson

English learners (EL) are one the largest population groups within the public schools in United States. Their academic achievement is lower than their monolingual peers in several areas. Working memory (WM), a cognitive process, has been identified as one of the predictors of poor academic achievement in both monolingual English speaking and EL children. Children at risk for learning problems have poor WM capacity. In addition, teachers have been acknowledged as one of the primary sources for identifying children at risk for learning problems. This study investigated the relationship between teachers' ratings of ELs' WM and students' actual WM performance. Because WM is composed of several components, a determination was made as to whether teacher ratings are related to isolated components of WM or whether ratings are predictive of all components. Because teacher ratings are context specific (student make-up of classroom), multilevel modeling was used to identify whether teacher ratings accurately predict EL children's performance on WM tasks. Additionally, covariates related to student characteristics and achievement were entered in the model to determine whether unique variance related to teacher ratings was sustained in predictions of WM performance. The results suggest that teacher ratings do predict both general WM and isolated components of WM in both English and Spanish. However, this relationship was not sustained when covariates related to student achievement were entered into the multilevel regression model. The results suggest that teacher ratings of low WM were more related to vocabulary and academic achievement than WM limitations.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I - INTRODUCTION

Educating the increasing percentage of English learners (ELs) in public schools across the United States has been a growing concern. Children who speak English as a second language are the fastest-growing population in United States. According to the National Center of Education Statistics, in 2011, approximately nine percent (4.7 million) of public school students were identified as English learners (ELs). Compared to Caucasian children and other minorities, academic achievement is lower for ELs especially those who speak Spanish (Lesaux, Crosson, Kieffer, & Pierce, 2010). In grades two through eleven, ELs have the lowest test scores than their non-EL peers, and account for two-thirds of the achievement gap California experiences in reading and mathematics performance compared to other states (Betts, Rueben, & Danenberg, 2000). In July 2013, California reformed the way schools are to be funded by passing the Local Control Funding Formula (LCFF) and Local Control Accountability Plan (LCAP). School districts now have to provide additional financial resources to support education of students with additional academic needs such as ELs. Districts will also be held accountable for what services they provide to support learning of students with needs.

Research has found ELs have disproportionately lower achievement compared to their monolingual peers, particularly in the area of reading. Since reading is a fundamental skill crucial to the achievement of ELs and their future success in the United States, there is much attention around exploring the cognitive mechanism that may underlie these reading difficulties. Some studies have explored the cognitive processes underlying reading problems in native English-speakers and ELs, and have indicated that working memory (WM) may play an important role in predicting reading performance (Geva & Siegel, 2000; Siegel & Ryan, 1989; Swanson, Orosco, & Lussier, 2011; Swanson, Sáez, & Gerber, 2006). Research has shown that children with low scores on measures of WM typically perform relatively poor on achievement measures (Gathercole & Alloway, 2008; Gathercole, Alloway, Willis, & Adams, 2006; Lesaux & Siegel, 2003; Swanson, Sáez, Gerber, & Leafstedt, 2004; Swanson, Orosco, & Lussier, 2011). The children with poor WM capacity frequently fail in classroom activities that involve following instruction and engaging in cognitively demanding activities (Gathercole, Lamont & Alloway, 2006). In addition, some studies observed the laboratory analogues of classroom activities under controlled conditions and confirmed that children with low WM scores have difficulties in classroom activities with high WM demands (e.g., Gathercole, Durling, Evans, Jeffcock, & Stone, 2008).

Working memory refers to the ability to hold and manipulate information for short periods of time. WM is considered a system of inter-linked memory components (Gathercole & Alloway, 2008). One of the primary models used to explore the role of WM in reading is Baddeley's multi-component model (Baddeley, 1986; Baddeley & Logie, 1999). This multi-component model is comprised of three WM components: a phonological loop, a visual-spatial sketchpad, and a central executive system that controls the other two subsystems. The two subsystems are usually referred to as short-term memory (STM) components. The phonological loop (verbal STM) temporarily stores verbal information. The phonological loop consists of two subcomponents, a store that holds phonological information and retrieval of information through rehearsal process

(Baddeley, 2006). The visual-sketchpad (visual-spatial STM) is responsible for storing non-verbal (visual-spatial) information for a short duration. The central executive system is responsible for directing attention to relevant information and controlling the subsystems. Research has identified strong links between various components of WM and academic achievement (Gathercole et al., 2006; Pickering & Gathercole, 2004; Swanson & Berninger, 1995).

Thus, given the importance of working memory in predictions of academic achievement, it is of value to understand teachers' awareness of working memory and whether they can accurately identify working memory deficits in children. This may be particularly difficult to identify in English learners due to a lack in appropriate assessments to separate out limited language proficiency from learning problems (Wagner, Francis, & Morris, 2005).

Statement of the Problem

Children with low working memory frequently fail to follow instructions, have problems with activities that combine storage and processing, and have difficulty keeping track of their progress in a particular activity (Gathercole & Alloway, 2008). Many students with poor working memory go undiagnosed and are instead labeled by teachers as being inattentive, highly distractible, and having short attention spans (Gathercole, Alloway, Kirkwood, Elliott, Holmes, & Hilton, 2008). Children with poor working memory and low achievements scores often receive higher ratings of inattention by their teachers (Gathercole et al., 2008).

In order to close the academic achievement gaps, early identification of children with working memory problems is extremely important. By identifying the children with working memory problems, better interventions can be provided which will enable them to learn. Research indicates that teachers' awareness of working memory deficits in the classroom is low, and teachers are only able to identify early signs of working memory problems in their students for 25% of the time (Gathercole et al., 2006). Teachers could only correctly identify one or two signs of WM failure in students and children that teachers identified as troublesome showed signs of WM failure (Alloway, Doherty-Sneddon & Forbes, 2012). However, there is limited research on teachers' awareness of working memory problems for EL children.

Purpose of the Study

Some studies have examined the effectiveness of Working Memory Rating Scale (WMRS, Alloway, Gathercole, & Kirkwood, 2008), and found WMRS to be a reliable screening tool which enables teachers to identify working memory problems in their students (Alloway, Gathercole, & Kirkwood, 2008; Alloway, Gathercole, Kirkwood & Elliot, 2009b; Pimperton & Nation, 2014). However, these studies have several limitations to this research. First, classroom observations have not been validated on WM measures when the context of these observations has been partialed out in the analysis. Teacher observations are nested within a classroom, and no studies to my knowledge has accounted for this variance (random effect). Second, teacher observations of WM may be influenced by several factors, such as reading achievement, classroom disruptions, and language proficiency and therefore such observations may not reflect actual WM performance. Studies have only explored the relationship between teacher observations of WM and children's' WM performance without accounting for factors that might influence this relationship. Finally, teacher observations of WM are viewed as reflecting a single general factor, but may instead reflect multiple components.

The purpose of the study is to investigate whether teachers, using the WMRS, can accurately identify working memory problems in EL children. Unlike their English monolingual peers, EL students have limited language proficiency in English, which creates a barrier for early identification of learning difficulties for this population. Furthermore, the study is an attempt to explain variation in teacher ratings that might be due to student achievement and characteristics.

Thus, this study addresses three questions:

1. Do teacher ratings predict performance on laboratory measures of WM and do these predictions reflect the relationship between a general WM system or specific components of WM?
2. Are teacher ratings of WM confounded by student characteristics related to achievement, language proficiency, gender, and grade?
3. Do teachers who accurately identify WM problems differ from teachers who incorrectly identify WM problems on student measures of English and Spanish reading, English and Spanish oral language, mathematics achievement, and age?

CHAPTER II – REVIEW OF THE LITERATURE

This review will discuss the importance of working memory in elementary school-age monolinguals and English learners. It will also detail the importance of early identification of learning difficulties in children and the role teachers play in identifying those problems. The gap in literature regarding teachers' awareness of working memory deficits in children, especially English learners will be reviewed.

English Learners

English learners (ELs) are students whose primary language is other than English and who have limited skills in the English language. Saunders and Marcelleti (2013) defined ELs as students whose families speak a language other than English and who scored as limited English proficient on the language proficiency assessments. In California, students whose parents respond to the Home Language Survey and indicate language other than English being spoken in the home are tested in schools within 30 days of enrollment for their English language proficiency using the California English Language Development Test (CELDT). CELDT assesses students on listening, speaking, reading, and writing in English (California Department of Education, 2015). There are five levels of CELDT scores (1) Beginner, (2) Early Intermediate, (3) Intermediate, (4) Early Advanced, and (5) Advanced. Based on their CELDT scores, students are identified as ELs if their overall proficiency falls below early advanced (Saunders & Marcelleti, 2003).

ELs are the fastest-growing population in the U.S. schools, with 10 percent increase annually (U.S. Department of Commerce, 2004, as cited in McCardle, Mele-

McCarthy, Cutting, Leos, & D'Emilio, 2005). In academic year 2010-2011, 4.7 million ELs were enrolled in the U.S. public schools. In year 2009-10, 29% of the enrolled students in California were ELs (U.S. Dept. of Education [U.S. DOE], 2013). Majority of the ELs (80%) in the United States are Hispanic/Latino students who speak Spanish (Aud, Hussar, Kena, Bianco, Frohlich, Kemp, & Tahan, 2011; Hemphill & Vanneman, 2011). ELs have the lowest achievement scores, highest dropout rates, and largest mobility rate than other student groups (McCardle et al., 2005).

Academic Achievement in ELs

Academic achievement is lower for ELs especially those who speak Spanish as their first language (Lesaux, Crosson, Kieffer, & Pierce, 2010). Latino ELs have the lowest average achievement scores on the national assessment of mathematics and reading (Fry, 2007). On the 4th grade standardized reading assessment of NAEP, the achievement gap between non-ELL and ELL students was 36 points, and 44 points at the 8th grade level (Aud, Wilkinson-Flicker, Kristapovich, Rathbun, Wang, & Zhang, 2013).

ELs have poor reading outcomes especially underdeveloped vocabulary and comprehension skills, and consistently lag behind their native English-speaking peers (August & Hakuta, 1997; August & Shanahan, 2006; Geva & Massey-Garrison, 2013). According to Betts, Rueben, and Danenberg (2000), ELs account for two-thirds of the achievement gap that California experiences in reading and math performance compared to other states.

Working Memory

One important predictor of academic achievement in children found in recent studies of both monolingual and ELs is working memory (e.g., Alloway, Gathercole, Adams, Willis, Eaglen, & Lamont, 2005; Berg, 2008; Engle, Carrullo, & Collins, 1991; Gathercole, Alloway, Willis, & Adams, 2006; Swanson, Orosco, Lussier, Gerber, & Guzman-Orth, 2011). Working memory is made up of several components (Baddeley & Logie, 1999). These components consist of a domain-general memory component that controls attention and involved in higher-level mental processes called central executive, and two domain specific slave systems known as phonological loop and visual-spatial sketchpad. Each of these components is briefly reviewed.

WM Components

Phonological loop. This is a system that stores material that can be expressed in spoken language such as numbers, words and sentences. It has two basic components, a phonological store and an articulatory rehearsal mechanism. The former is simply responsible for holding the information temporarily, and the latter is used to prevent the rapid decay by reciting the information in the phonological store.

Visual-spatial sketchpad. This is a system for storing visual and spatial information such as images and locations. It is responsible for recalling shapes, visual representations of objects and movements.

Central executive control system. Central executive system controls attentional processes rather than act as a storage. It is responsible for “coordinating the two slave systems, focusing and switching attention, and activating representations within long term memory" (Baddeley & Logie, 1999, p. 28).

Gathercole and Pickering (2000) developed a battery of tests guided by Baddeley and Hitch's (1974) WM model to analyze the function of WM in six- and seven-year old children. Their study found support for the phonological loop tasks that involved temporary verbal storage and central executive tasks that measure both processing and storage of information. However, little evidence was found regarding visual spatial and whether it taps a distinct component of WM. In another study, Gathercole, Pickering, Ambridge, and Wearing (2004) tested the structure of WM in children four- to 15- years of age and used multiple assessments for each component of WM model. Confirmatory factor analysis supported the tripartite model. The findings suggest that the structure of WM system is in place by 6 years of age and the remains stable as children develop.

Alloway, Pickering and Gathercole (2006) further investigated the structure of verbal and visuospatial short-term and WM in children and found support for a three-factor model. They found that the three-factors are related but separable factors that measure verbal and visuospatial storage. The third factor accounted for the relationship between STM and WM measures and represented the shared variance between the verbal and visuospatial WM tasks.

Working memory is crucial for learning and academic success. WM plays a critical role in simple skills such as decoding (Kail & Hall, 2001) and arithmetic calculation (Berg, 2008; DeStefano & LeFevre, 2010), as well as in higher level cognitive skills as reading comprehension (Daneman & Carpenter, 1980; Engle, Carullo, & Collins,

1991) and word problem solving (Swanson & Beebe-Frankenberger, 2004; Zheng, Swanson, & Marcoulides, 2011). Reliable associations have been found between WM skills and children's scores on national curriculum assessments in both English and mathematics by the age of 7, and with achievement in mathematics and science by 14 years of age (Gathercole, Pickering, Knight, & Stegmann, 2004). While WM is an important predictor in many areas of academics, the focus of the review will be on the role of WM in reading.

Role of WM in Reading

Working memory has been linked with early reading achievement in both monolingual English speakers and EL children (Gathercole et al., 2004; Gottardo, Stanovich & Siegel, 1996; Lesaux, Lipka, & Siegel, 2006; Siegel & Ryan, 1989; Swanson & Berninger, 1995). While working memory is an interlinked system of three components, phonological loop, central executive system, and visuo-spatial sketchpad, different activities engage in some or all of the WM components (Gathercole & Alloway, 2008).

Reading acquisition requires word decoding and reading comprehension. Word decoding refers to one's ability to sound out words and read isolated words quickly, silently, and accurately (Gough & Tunmer, 1986). Reading comprehension includes several skills such as decoding, assembling word meaning into larger units, constructing meaning of sentences and linking information across sentences, focusing attention on main ideas, and integrating information (Dehn, 2008).

Previous research has established that STM is more likely to predict word identification (Chiappe, Siegel, & Wade-Woolley, 2002; Jongejan, Verhoeven, & Siegel, 2007), and the process of comprehension requires both verbal STM and WM (Alloway, Gathercole, Willis, & Adams, 2004; Daneman & Carpenter, 1980; Linck, Osthus, Koeth, & Bunting, 2013). In order to comprehend the text, a reader not only maintains the words of a sentence in memory (verbal STM) but also integrates information through text, establishes links between sentences, and draws on general knowledge, which involves WM (Cain, 2006). WM has also found to contribute unique variance to reading comprehension above vocabulary (Cain, Oakhill, & Bryant, 2000).

In earlier work, Daneman and Carpenter (1980) studied the role of WM in comprehension. They developed a task called WM span, which required both processing of sentences and remembering the last word of each sentence. High correlation between the WM span task and measures of reading comprehension were found in their sample of college students. While correlation does not prove causality, many studies have replicated the findings with children as well.

Similar to monolingual students, working memory is an important indicator of reading success for EL students (Engel de Abreu & Gathercole, 2012; Swanson, Sáez, Gerber, Leafstedt, 2004; Swanson, Orosco, Lussier, Gerber, & Guzman-Orth, 2011). Swanson et al. (2004) examined cognitive processes that underlie second language (L2) reading difficulties for first-grade children whose first language is Spanish, and discovered that while STM is important for L2 reading acquisition, the executive process of WM also significantly relates to L2 reading and vocabulary (Swanson et al., 2004).

Spanish and English WM contributed important variance to English word reading and vocabulary that is distinct from STM. In their longitudinal study, Swanson, Sáez, and Gerber (2006) discovered that growth in WM in children's primary language (Spanish) predicted the growth in their second-language (English) reading.

Swanson et al. (2011) further examined the contribution of working memory components to children's L2 reading and language acquisition, and found both STM and WM contributed unique variance to L2 reading. English word reading and vocabulary were predicted by phonological loop (STM) and the executive component of WM. Visual-spatial sketchpad was also found to predict reading comprehension. In their study, children were given cognitive and reading measures in both English and Spanish. Weak cross-language transfer was found; English measures were found to best predict English literacy and Spanish measures best predict Spanish literacy.

WM in the Classroom

WM plays an important role in the classroom. Certain classroom situations cause difficulties for children with poor WM. Children's inability to meet the WM demands of the classroom activities has been discussed as the "working memory overload hypothesis" by Gathercole et al. (2006). Due to the WM overload, the children showed frequent task failures that further impaired their rates of learning (Gathercole et al., 2006).

Gathercole and Alloway (2008) observed several students with poor WM skills and noticed the memory-related failures that children experienced included difficulties in following multi-step instructions provided by teachers, failing to complete common classroom activities that required large amount of information to be held in mind, and problem keeping their place in demanding and complex activities such as writing. For example, tasks such as generating and writings a sentence requires letters, words, and the sentence, which place greater load on WM than just copying a sentence (Gathercole & Alloway, 2008; Gathercole, Alloway, Kirkwood, Elliot, Holmes, & Hilton, 2008; Gathercole, Lamont, & Alloway, 2006).

Children with WM deficits also demonstrated having high levels of inattentive and distractible behavior (Gathercole & Alloway, 2008).

WM and Inattentive Behavior

WM impairments have been linked to attentional problems (Gathercole & Alloway, 2008; Kofler, Rapport, Bolden, Sarver, & Raiker, 2010; Rapport Bolden, Kofler, Sarver, Raiker, & Alderson, 2009). Gathercole, Durling, Evans, Jeffcock, and Stone (2008) investigated whether poor working memory is accompanied by attentional difficulties in children, and found that teacher rated students with low WM scores as having high levels of problem behaviors related to cognitive problems and inattention.

In the same study, five-and six-year old children were given spoken instructions regarding manipulation of a sequence of objects, and they were asked to either perform the task based on instructions of repeat them. Children with low WM scores struggled to perform verbal instructions. Similarly, Alloway, Gathercole, Kirkwood, & Elliot (2009b) found that children with low WM were rated having poor attention span and high levels of distractibility. Children with WM impairments often fail in the classroom because the

WM loads of each activity exceed their capacities. Due to this, children forget what they are doing, and this leads to inattentive behavior.

Studies have investigated the functional relationship between WM and inattentive behavior in children with ADHD and typically developing children (Alloway, Gathercole, Holmes, Place, Elliot, & Hilton, 2009; Kofler et al., 2010; Rapport et al., 2009), and the findings suggest that attention behaviors associated with poor WM are different from behaviors related with ADHD. Children with ADHD had impulsive and disruptive behaviors; whereas, children with poor WM were more inattentive and distracted (Gathercole & Alloway, 2008). In Alloway et al.'s study (2009b) teachers rated children with ADHD more highly in oppositional and hyperactive behaviors, and children with WM impairments rated high on WM difficulties behaviors that included planning and organizing information. The findings of St Clair-Thompson's study (2011) also revealed that children with poor WM had significantly lower scores on measures of planning and attention.

Children with WM deficits have also been linked to at risk for learning difficulties (e.g., Gathercole et al., 2006; Jerman, Reynolds, & Swanson, 2012; Siegel & Ryan, 1989; Swanson et al., 2004; Swanson et al., 2011).

WM and Learning Difficulties

Children with learning difficulties in reading and math typically have poor WM capacities, and their memory scores predict the severity of their learning problem (e.g., Alloway, Gathercole, Kirkwood, Elliot, 2009a; Mih & Mih, 2011). Relationship between WM capacity and learning difficulties has been found in EL children at risk for learning difficulties (Swanson et al., 2011). Children at risk for learning problems with WM deficits are unable to retain information in memory while at the same time process same or other information (Swanson & Zheng, 2013).

Research has consistently found children with reading or math problems experience difficulties in WM tasks (Gathercole, Alloway, Willis, & Adams, 2006; Swanson & Berninger, 1995). The review of memory difficulties in children with learning difficulties suggests that these children may experience deficits in executive system, and those with reading difficulties experience problems in phonological processing (Swanson & Zheng, 2013).

Pickering and Gathercole (2004) investigated WM skills in 98 children, ages 4 to 15 years, with special educational needs and discovered that children performed poorly on measures of all three components of WM. In their earlier study, Gathercole and Pickering (2001) found that deficits in WM were related to poor performance in reading, writing, spelling, and mathematics for young children. Findings from studies that compared children with RD to typically achieving peers suggest that children with RD had weaker verbal WM spans (Siegel & Ryan, 1989; Swanson & Howell, 2001; Swanson, Mink, & Bocian, 1999).

WM deficits can lead to lost learning opportunities and slow rates of educational progress (Alloway, 2006; Gathercole & Alloway, 2008). Due to the strong relationship between WM deficits and LD, Dehn (2008) suggests that WM should be assessed when a child is referred for a possible LD.

Early Identification of Learning Difficulties

Research has shown that 10 to 15 percent of children in mainstream classrooms will suffer from WM impairments that will jeopardize their academic success (Alloway, 2010). Therefore, early identification of poor working memory skills is clearly needed because of its relationship to academic success (Dehn, 2008). Due to the strong link between WM and learning difficulties, it is important to review literature regarding challenges around early identification of learning difficulties.

Steele (2004) makes the case for early identification and intervention for young children who are at risk for learning difficulties. Steele argues that early identification and intervention is important because not only it is foundation for later learning, but it can also prevent secondary problems from occurring. Additionally, children who are identified early enough are likely to have the opportunity to develop to their potential. While early identification has been a challenge for all children with difficulties, it is particularly a problem when identifying learning difficulties in children who are learning in a second language.

Early Identification of Learning Difficulties for ELs

Appropriate and timely identification of EL at risk for academic difficulties has been an ongoing challenge for educators in U.S. schools (Artiles & Ortiz, 2002). For EL children, it is “unclear whether limited language proficiency in English is interfering with learning or is masking a learning disability, or leads to poor performance on assessments used for identification, which are not culturally and linguistically appropriate for that purpose” (Wagner, Francis, & Morris, 2005, p. 6). This makes it difficult to identify EL children who have learning problems or at risk for learning disability. Schools fail to identify due to the inability to distinguish language proficiency issue from learning difficulties. Often, EL children are identified for special education services after continuously underperforming for several years (McCardle et al., 2005).

Samson and Lesaux (2009) investigated the proportional representation, identification rates, and predictors of ELs in special education in kindergarten, first grade, and third grade. They found that ELs are underrepresented in special education in kindergarten and first grade; however, they are overrepresented in third grade in certain categories, including LD and speech-language impairment (Artiles & Ortiz, 2002; Samson & Lesaux, 2009).

Limbos and Geva (2001) discovered that EL children are less likely to be identified as at risk compared to their monolingual English speaking peers. ELs were overlooked for services for reading difficulties due to the lack of language proficiency (Limbos & Geva, 2001; Samson & Lesaux, 2009). Even when language proficiency is not in question, there is still a delay in identifying children with learning difficulties, and most are not identified until second or third grade (Wagner, Francis, & Morris, 2005). Reading problems becomes well-rooted and more difficult to overcome the longer they exist. Therefore, early identification can lead to teachers providing targeted interventions for children with reading difficulties.

Testing ELs in their native language or in English has also been a growing concern (Wagner, Francis, & Morris, 2005, p. 10). Therefore, teacher judgments play an important role in identifying children with special service needs.

Role of Teachers in Early Identification

Teachers may be one of the best resource for predicting children who are at risk for learning difficulties (e.g., Salvesen & Undheim, 1994; Samson & Lesaux, 2009; Sudkamp, Kaiser, & Moller, 2012). They are a primary link in identifying at-risk students and delivering services to those identified students. Salvesen and Undheim (1994) investigated the use of teacher assessments in screening for LD. Over 600 children were rated by their teachers in the second grade, and teachers were found to be mostly accurate in their identifying children with low achievement; however, they were less accurate in their identification of specific reading difficulties.

In other studies, teacher predictions have been found to be accurate in the process of early identification of reading and/or learning difficulties (Samson & Lesaux, 2009; Steele, 2004). When compared to children's reading proficiency or EL status, teacher ratings were stronger predictors of placing children in special education (Samson & Lesaux, 2009).

Speece, Ritchey, Silverman, Schatshneider, Walker, and Andrusik (2010) found that teacher ratings of reading problems are a significant predictor of at-risk status in fourth grade children. Therefore, teacher ratings of academic progress provide a promising alternative to existing methods of identification of learning problems (Taylor, Anselmo, Foreman, Schatschneider, & Angelopoulos, 2000). Some studies looked at teacher predictors as early as kindergarten and established that teacher ratings of literacy skills was highly predictive of early identification of learning problems and children's placement in special education (Taylor et al., 2000).

It is evident that teachers play an important role in early identification of children with LD. However, Sideridis, Antonious, and Padeliadu (2008) tested for teacher biases in the identification of learning disabilities and discovered that gender of teacher was associated with biases, meaning more children who had LD were identified by female teachers than male teachers.

Teacher Knowledge of WM

Studies suggest that teachers' awareness of WM is limited. Teacher interviews revealed that teachers never listed WM as an explanation for students' experiencing difficulties (Alloway, Doherty-Sneddon, & Forbes, 2012). Teachers in Alloway, Doherty-Sneddon, and Forbes's (2012) study could only correctly identify one or two signs of WM failure in their students, and rated the students with WM deficits as troublesome. Teachers gave higher ratings of inattention for children with lower achievement and memory span scores (Gathercole et al., 2008).

Based on the teacher interviews, Alloway and colleagues developed a Working Memory Rating Scale (WMRS) for teachers consisting of a checklist based on key behavioral characteristics that children with WM impairments exhibit. WMRS is considered a more accurate assessment of problem behaviors associated with WM difficulties compared to Conners' Rating Scale (CRS-R) and Behavior Rating Inventory of Executive Function (BRIEF) (Alloway et al., 2008; Pimperton & Nation, 2014). The CRS-R is a teacher rating scale used in assessing children and adolescents with ADHD, whereas, BRIEF is a teacher rating scale that assesses the executive function and self-regulation in children and adolescents.

Alloway et al. (2009) compared the teacher ratings on the WMRS with the CTRS and BRIEF in low WM children. The WMRS was found to identify greater proportion of children with WM deficits than using the CTRS and BRIEF. Pimperton and Nation (2014) study used the WMRS, and found that children with poor comprehension skills were rated high on the WMRS compared to the control group.

Alloway, Gathercole, Kirkwood, and Elliot (2009b) investigated the psychometric properties of the Working Memory Rating Scale (WMRS) on over 400 elementary school children from schools in North-East England. The WMRS had a good internal reliability and convergent validity with measures of WM. The authors identified WMRS as a useful screening tool for teachers to detect children with WM difficulties. However, other studies that investigated the psychometric properties with different samples found different results.

Normand and Tannock (2014) studied the psychometric properties of the WMRS in 524 six- to nine-year old children in Canada, and found a poor fit for the 20 items on the WMRS. Instead they found a support for a five-item short form. However, the authors did identify WMRS as a useful and time-effective tool for teachers to screen WM deficits in their children.

The majority of the research on teacher identification of WM deficits in children has been conducted with samples from monolingual English speaking children in countries outside of United States. Guzman-Orth, Grimm, Gerber, Orosco, Swanson, and Lussier (2014) study is the only study that tested the reliability and validity of WMRS on a sample of Spanish-speaking ELs. Unlike the findings of Normand and Tannock (2014) study, Guzman-Orth et al.'s (2014) study found a one factor fit the data. Similar to Alloway et al. (2008), convergent validity was established with the CTRS, a behavioral rating measure. Findings from this study were dissimilar to Alloway et al. (2009b) findings since a weak relationship was found between the WMRS and the measures of WM.

Alloway et al. (2009b) identified two major challenges to effectively identifying WM problems in children in the classroom, "first, working memory problems are difficult to detect from casual observation alone, and secondly, that there is an absence of suitable assessment tools that can be used by teachers to identify potential working memory problems" (p. 245). More research is warranted to establish WMRS as a suitable assessment for EL children in United States schools and whether teachers rate their ELs using WMRS are accurate in their judgments.

Statement of Purpose

Research on teachers' accurate identification of WM problems in children has been limited to monolingual English students. While some studies on EL children have looked at teacher ratings of WM as a predictor of academic achievement, to best of my knowledge, no study, however, has looked at the accuracy of teacher ratings independent of measures of achievement and language proficiency. I plan to extend the literature by examining the accuracy of teacher ratings of WM as a function student characteristics related to achievement, behavior and language proficiency.

Research Questions

1. Do teacher ratings predict performance on laboratory measures of WM and do these predictions relate to a general WM system or specific components of WM?
2. Are teacher ratings of WM confounded by student characteristics related to achievement, language proficiency, gender, and grade?
3. Do teachers who accurately identify WM problems differ from teachers who incorrectly identify WM problems on student measures of English and Spanish reading, English and Spanish oral language, mathematics achievement, and age?

CHAPTER III - METHODOLOGY

This study used a secondary dataset from a grant funded by the Institute of Educational Sciences entitled Growth in Literacy, Language and Cognition in Children with Reading Difficulties who are English Language Learners (R324A090092). The data was collected over a three year period from 2009 to 2012; however, this study only used the first year data. The purpose of the grant was to identify the measures and processes that accurately identify children at risk for reading disabilities who are English learners.

Participants

Data was collected from elementary school children from three large school districts in the southwestern United States. All children in the study were Hispanic and the first language for all children participating in the study was Spanish. Majority of the children (80%) spoke Spanish in the home, 8% spoke both English and Spanish, and 10% spoke only English in the home. Children were designated as EL based on California English Language Development Test (CELDT) administered in the school. Their scores ranged from (1) beginning, (2) early intermediate, (3) intermediate, (4) early advanced, and (5) advanced. Children received classroom reading instruction in either English or a combination of English and Spanish. Ninety-seven (97%) percent of the sample participated in a federally funded free lunch program. In the first year of the study, a total of 500 students were tested. The sample included 234 male and 266 female children. Of the 500, 163 students were in the first grade, 153 in the second grade, and 184 in the third grade.

Measures

Teacher Rating Scales

Two behavior rating scales, the Working Memory Rating Scale and the Conner's Teacher Rating Scale, were completed after testing was concluded in the spring by each of the participating student's teachers.

Working Memory Rating Scale, Teacher Version (WMRS). The WMRS is a behavioral rating scale developed for teachers to aid easy identification of children with working memory deficits (Alloway, Gathercole, & Kirkwood, 2008). Teachers rarely identify memory as a source of difficulty in children with working memory problems; instead, children with memory problems are typically described as inattentive. The WMRS is aimed at increasing the chances of the detection in children with deficits of working memory. The WMRS was developed on the basis of interviews with teachers and consists of 20 items. The items include: “The child raised his hand but when called upon, he had forgotten his response”; or “She lost her place in a task with multiple steps”. Teachers rate how typical each behavior was of a child, using a four-point scale ranging from (0) not typical at all to (1) occasionally to (2) fairly typical to (3) very typical. The WMRS provides a raw score that is then converted to a T-score. A T -score is a type of standardized score with a mean of 50 (SD=10). T-scores have an average range of 40 to 60 and 68% of the population would score within that range.

Conners’ Rating Scale – Revised (CTRS). The Conners' Teacher Rating Scale (Conners) is a commonly used instrument by the teachers to assess children's behavior in the classroom (Conners, 1997). It consists of 28 items that represent internalizing and externalizing behaviors on four subscales (Oppositional Scale [negative/defiant behaviors], Cognitive Problems/Inattention, Hyperactivity, and Attention-Deficit/Hyperactivity Disorder [ADHD]). Only the Cognitive Problem/Inattention subscale was used in this study because children with low WM ability have low attention span and are highly distractible (Alloway et al., 2009b; Gathercole & Alloway, 2008; Gathercole et al., 2008; St. Clair-Thompson, 2011). The other subscales of CTRS are related to hyperactive, opposition behaviors that are found more in children with ADHD than in children with WM impairments (Alloway et al., 2009b).

Children WM Measures

Previous studies have shown that different memory measures load on to different components of WM (e.g., STM, executive WM, and Visual-Spatial WM; Swanson et al., 2011; Swanson, Orosco & Lussier, 2011; Swanson, Orosco, & Lussier, 2015). The Forward and Backward Digit Span, Word Span, and Pseudoword Span tasks load on the phonological loop component; whereas, four measures that required children to respond to a processing question in addition to recall (i.e., Conceptual Span, Listening Sentence Span, Rhyming Span, and Updating) tapped on the executive component of WM. Visual-Matrix and Mapping & Direction tasks loaded on Visual-spatial sketchpad.

Phonological loop (Short-term memory; STM). Four measures of short-term memory (STM) were administered in Spanish and English: Forward Digit Span, Backward Digit Span, Phonetic Memory Span, and Word Span.

Forward and backward digit span. The Forward and Backward Digit Span tasks (taken from the WISC-III; Wechsler, 1991) were administered along with experimental

Spanish translated versions. For the Forward Digit Span task, children had to recall sequentially ordered sets of digits that increased in number, which were spoken aloud by the examiner. The Backward Digit Span task required children to recall sets of digits, but in reverse order. Dependent measures for both tasks were the largest set of items recalled in order.

Phonetic memory (pseudoword) span task. The Phonetic Memory Span (adapted from a task in Swanson & Berninger, 1995 and Swanson & Beebe-Frankenberger, 2004) is an experimental task designed to assess phonological short-term memory for nonwords (a.k.a. pseudo-words) in an English and Spanish version. Children had to recall a list of nonwords (one syllable long) that they were told by the examiner. There was a gradual increase in set size, up to eight words in the list. The dependent measure was the highest set of items retrieved in the correct serial order (range of 0 to 6).

Word span task. The task (a.k.a. Real Word Span, task adapted from Swanson, Ashbaker, & Lee, 1996; and Swanson & Beebe-Frankenberger, 2004) assessed short-term memory retention for real words in an English and Spanish version. This was also an experimentally-designed memory task, for which, examiners orally presented a list of common, but unrelated nouns that they were asked to recall. Word lists gradually increased in set size from a minimum of two words to a maximum of eight. Similar to other STM tasks, the dependent measure was the highest set of items recalled.

Executive system of WM. A Conceptual Span, Listening Sentence Span, Rhyming Words, and Updating tasks were administered in English and Spanish to capture the executive component of WM in both language systems. The executive component WM tasks are different from the phonological span tasks because they required children to hold increasingly complex information in memory while simultaneously responding to a discrimination question.

Conceptual span. The Conceptual Span task assessed the students' ability to organize sequences of words into abstract categories (adapted from the Semantic Association subtest of the S-CPT; Swanson, 1996b, and Swanson, 2008). In this task, children listened to a set of words that, when re-organized, could be grouped into meaningful categories. They had to retrieve the words that "go together", and simultaneously answer a discriminating question. The task required students to transform information encoded serially into categories during the retrieval phase. The dependent measure was the highest set recalled correctly (range of 0 to 8) and in which the discriminating process question was also answered correctly. A Spanish version was created and administered to the children.

Listening sentence span task. This task is an experimental children's adaptation (Swanson, 1992; 1996a; 1996b) of Daneman and Carpenter's (1980) Listening Sentence Span Task. A Spanish version was created and administered to the children. The children were presented with groups of sentence read aloud, and they had to simultaneously understand the sentence contents and remember the last word of each sentence. The number of sentences in the group gradually increased from 2 to 6. After each group of sentences was presented, the child answered a discriminating process question about a sentence and then was asked to recall the last word of each sentence in order. The dependent measure was the total number of correctly recalled word items in order up to the largest set of items, in which the process question was also answered correctly.

Rhyming words. The Rhyming Span subtest from the Swanson Cognitive Processing Test (S-CPT) and the adapted experimental Spanish version (S-CPT; Swanson, 1996b; Swanson, Howard, Sáez, 2006) was administered to assess children's processing of acoustically similar words in English and Spanish. Students were presented with a series of rhyming words, asked a discriminating process question, and were then asked to recall the words in order. After answering the process question correctly, the students had to repeat the sequence in order. The word sets gradually increased in length; the difficulty ranged from a set of two rhyming words to a set of fourteen rhyming words. The dependent measure for both versions was the number of sets recalled correctly. A parallel Spanish version was also administered (Swanson & Beebe-Frankenberger, 2004).

Updating task. An experimental Updating task adapted from Swanson, Sáez, Gerber, & Leafstedt (2004) was administered in both English and Spanish. For this task, a series of one-digit numbers were presented that varied in set lengths of three, five, seven, and nine. No digit appeared twice in the same set. The examiner told the child that the length of each list of numbers might be three, five, seven, or nine digits. Students were then told that they should only recall the last three numbers presented in the set. Each digit in the list was presented at approximately one second intervals. After the last digit was presented the participant was asked to name the last three digits in order. The dependent measure was the total number of sets correctly repeated (range 0 to 16).

Visual-spatial sketchpad of WM. Mapping and Directions Span and Visual Matrix tasks were administered to assess the visual-spatial sketchpad of the WM system.

Mapping and direction span test. The Mapping and Direction Span subtest from the Swanson Cognitive Processing Test (S-CPT; Swanson, 1996b; 2011) was used to determine whether the students could recall a visual-spatial sequence of directions on a map with no labels (Swanson, 1996b; 2011). Children were presented with a street map for ten seconds with lines connected to a number of dots. Prior to presentation, the child was shown a card with four strategies for encoding visual-spatial information for later recalled. After removal of the map, the child was asked, "Are there any stoplights in the first column?" and asked to circle "Y" for yes or "N" for no. Then, the child was asked what strategy s/he would use to remember the street arrows and stop lights. The child was instructed to draw the directions and stoplights on a blank map.

Visual matrix task. The Visual Matrix Task from the Swanson Cognitive Processing Test (S-CPT; Swanson, 1996b; 2011) was used to assess the working memory capacity of participants to remember visual sequences within a matrix (Swanson, 1996b; 2011). Children were presented a series of dots in a matrix (a grid made of squares) and were allowed five seconds to study the pattern. The matrix was then removed and participants were asked, "Are there any dots in the first column?" After answering this discriminating process question about the column (by circling "Y" for yes or "N" for no), students were then asked to draw the dots they remembered seeing in the corresponding boxes of a blank matrix in their response booklets. The task difficulty ranged from a matrix of four squares with two dots to a matrix of 45 squares with 12 dots. The dependent measure was the highest set recalled correctly (range of 0 to 11) in which the process question was answered correctly.

Achievement and Language Measures

Reading. Reading measures consisted of word identification and comprehension tasks which were administered in both English and Spanish.

***Letter-word identification*.** Real letter-word reading efficiency was assessed by the Letter-Word Identification subtest 3 from the Woodcock-Muñoz Language Survey-R (WMLS-R; Woodcock, Muñoz-Sandoval, & Alverado, 2005) in English and Spanish. Children were tested individually by presenting them with a list of words, which gradually increased in difficulty.

Passage comprehension. The vocabulary and short reading passage comprehension skills subtest from the Woodcock-Muñoz Language Survey-Revised (WMLS-R; Woodcock, Muñoz-Sandoval, & Alverado, 2005) was used to test comprehension. For this test, students had to fill in the blank spaces of various passages with specific words. The test was administered in both English and Spanish. The test yields a raw score that is converted to a standard score (M = 100, SD = 15). The internal reliability of the subtest is reported as .84.

Oral language. Children were measured on expressive vocabulary, receptive vocabulary, and syntax in both English and Spanish.

Expressive vocabulary. The Expressive One-Word Picture Vocabulary Test-Spanish and Bilingual edition (EOWPVT-SBE; Brownell, 2001) was used to measure the expressive vocabulary of children. Children were shown a test plate with a picture and asked to identify it. The EOWPVT-R was administered in both languages, with the first language randomly chosen for administration determining the order, until the child achieved a ceiling. If the child achieved a ceiling in English before Spanish, Spanish alone was continued and vice versa.

Receptive vocabulary. The Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981) was administered to measure children' receptive vocabulary in English, in which children were asked to select the picture from four options that matched the word read aloud by the tester. In Spanish, the Test de Vocaulario en Imagenes Peabody (TVIP; Dunn, Lugo, Padilla, & Dunn, 1986) was administered, which contained universally appropriate items from the PPVT-R.

Syntax. The Morphological Closure subtest from the Illinois Test of Psycholinguistic Ability III (ITPA-III; Hammill, Mather, & Roberts, 2001) was used to measure children's oral grammar skills. In this task, children had to provide a best fitting grammatically correct word for each sentence read aloud to them with a missing word. The items were translated into Spanish by three name speakers, and administered in the same manner as the English sentences.

Mathematics. The arithmetic subtest from the Wide Range Achievement Test (WRAT-3; Wilkinson, 1993) was administered to measure basic calculation ability. Children had to perform written computation on number problems that increased in difficulty. The subtest included single-digit addition items (2 + 2 = ?) to more advanced skills such as algebra. Children were allowed up to 15 minutes to complete math calculations. The dependent measure was the number of problems correct (raw score range was 15-55).

Procedures

Children were administered a battery of tests both in groups and individually across three waves. Battery of tests included measures of achievement, short-term-memory, executive system of working memory, and visual-spatial sketchpad of working memory. Most tests were administered in both in English and Spanish, and instructions were matched with the language of the test. Order of the language and measurements were counterbalanced. Children were also rated by their teachers on classroom behaviors and working memory.

Data Analysis

The current study used two types of statistical analyses to investigate the proposed research questions. For research question 1 and 2, multilevel regression modeling was used. For research question 3, generalized linear modeling was used to compare groups. The analyses were run using the SAS version 9.4 software program (SAS Institute, 2015).

Data Preparation

Prior to conducting any analyses, the distribution of the observations was checked. Means, standard deviations, skewness and kurtosis were examined for univariate normality. In addition, the data was screened for multivariate normality, outliers, and missing values.

Factor scores. To reduce the multiple measures to single variables, SAS CALIS program was used to create factor scores for each set of measures with two or more variables. Standardized beta weights were calculated, and based on the standardized loadings, factor scores were computed by multiplying the z score of the target variable by the standardized factor loading weights based on the total sample (see Nunnally & Bernstein, 1994, p. 508 for calculation procedures). Factor scores were created for English and Spanish *WM executive processing* (listening sentence span, and updating) English and Spanish *STM* (forward digit span, backward digit span, real word span, and pseudoword span), WM *visual-spatial sketchpad* (visual matrix, mapping/direction task), English and Spanish *reading* (word recognition, comprehension), and English and Spanish *oral language* (expressive vocabulary, receptive vocabulary, syntax).

Baddeley's WM model. The measurement models were examined to determine whether Baddeley's multicomponent WM model fit the data within each language. A confirmatory approach was taken because support has been found for Baddeley's model in children. Gathercole, Pickering, Ambridge, and Wearing (2004) examined the structure of WM and its development in children, and their findings suggest that the model of WM proposed by Baddeley is evident in children by 6 years of age. Similarly, Alloway, Pickering, and Gathercole (2006) investigated various WM models in children between ages 4 and 11 years, and found a support for a three-factor model that consists of two storage factors that are related but separable, and a single domain-general working memory factor that represented the shared variance between the verbal and visuospatial WM tasks.

WMRS factor structure. The primary predictor, teachers' ratings of WM, is a 20 item rating scale. Prior to analyzing the multilevel regression model, factor analytic technique will be used to determine the factor structure for the items. This helps determine whether teacher observations of WM reflect a single general factor or multiple factors.

Model Fit Criteria

Confirmatory factor analysis fit criteria. The chi-square goodness-of-fit test, the χ^2/df ratio, the comparative fit index (CFI), the Bentler-Bonett non-normed index (non-normed), and the root mean square error of approximation (RMSEA) were used to evaluate the model fit. A model is considered a good fit, meaning it adequately describes the sample data, if there is a consensus among the following criteria: a nonsignificant chi-square goodness-of-fit value, CFI and non-normed values greater than 0.90, and an RMSEA of .05 or less, along with having lower 90 percent confidence limit at or smaller than .05. Chi-square test is sensitive to sample size; therefore, more emphasis is placed on other fit criteria (Raykov & Marcoulides, 2008).

HLM models fit criteria. In order to assess the relative model fit and compare models, either a likelihood-ratio difference test is conducted or the information criteria associated with each model is compared. The likelihood-ratio difference test compares a more restricted model to the less restricted model, however, the difference test is only appropriate when one model is nested in the other. Information criteria, known as parsimony indices, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are also used to compare models. AIC is a measure of the goodness of fit of a model that considers the number of model parameters; whereas, BIC considers both the numbers of the parameters and the number of observations. These statistics are based on the value of -2 times the loglikelihood of the model, adjusted for the number of parameters in the model, and the sample size (only in BIC). The model with the fewer parameters is better; therefore, smaller AIC or BIC mean more parsimonious model.

Research Question 1 Analysis

Data was analyzed using a hierarchical linear modeling (HLM; also known as multilevel regression modeling; Raudenbush & Byrk, 2002) approach to take into consideration the nested effects, children nested in clasvsrooms. The fixed and random effect parameter estimates were obtained using PROC MIXED in SAS 9.4. The first step in the HLM is to run an unconditional model (Model 1 in Table 1) also known as one-way ANOVA with random effects. This model examined whether children's working memory scores varied across classrooms.

A sufficient sample size at level 2 is an important factor in multilevel analysis for accurate estimation. There were 79 classrooms in this study. Mass and Hox (2005) conducted a simulation study to determine the acceptable lower limit to the sample size and influence of sample sizes at the group level on the accuracy of the estimates. Their study found that sample size smaller than 50 at level two leads to biased estimates of the second-level standard errors, but the estimation of the regression coefficient, the variance components, and the standard errors remained unbiased and accurate in all the other simulated conditions. Thus, the present study meets the criteria for a minimum sample size.

In addition, the intra-class correlation (ICC) statistic was calculated to determine the proportion of the variance in the children' WM scores that is between classrooms, which helped justify the use of HLM. The unconditional model tested was:

Level 1 - $WM_{ij} = \beta_{0j} + r_{ij}$

Level 2 - $\beta_{0j} = \gamma_{00} + u_{0j}$

where *i* indexes students and j indexes classrooms.

In order to determine the effect of student-level predictor (teacher ratings of WM) on children' WM scores, a conditional model, random-coefficient model (Raudenbush & Bryk, 2002) was analyzed. In model 2, the teacher ratings of student WM was entered as a level 1-predictor and children' actual scores on WM was the dependent variable. The conditional model tested was:

Level 1 - $WM_{ij} = \beta_{0j} + \beta_{1j}(teacher\ WM\ rating) + r_{ij}$

Level 2 - $\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{10}$

Research Question 2 Analysis

The analysis for research question two was an extension to research question 1 where the conditional model was extended to include covariates. A conditional model (Model 3) with additional level 1 covariates was examined. Variables related to student achievement (reading, oral language, mathematics, and CELDT) and characteristics (gender and grade) were entered in the model. This model helped determine whether unique variance related to teacher ratings emerge in predictions of WM performance after controlling of these covariates. The variable CELDT (language proficiency) was included in the model in addition to oral language measures because CELDT is a state-wide measure of language proficiency. Additionally, CELDT is an overall English proficiency exam that takes into account reading, speaking, and writing skills. In the mixed-model form, the model tested was:

$$WM_{ij} = \gamma_{0j} + \gamma_{1j}(teacher\ WM\ rating) + \gamma_{2j}\ (E - reading)$$

$$+\ \gamma_{3j}(S - reading) + \ \gamma_{4j}(E - oral\ language) + \ \gamma_{5j}(S - oral\ language)$$

$$+ \gamma_{6j}(mathematics) + \gamma_{7j}(CELDT\_high) + \gamma_{8j}(gender\_male) + \gamma_{9j}(grade\_2)$$

$$+ \gamma_{10j}(grade\_3) + u_{0j} + r_{ij}$$

In order to examine whether teacher ratings are related to isolated components of WM, models 1, 2, and 3 were run for each WM component as a dependent variable. All the predictors in the models were grand-mean centered for ease of interpretation.

Test of assumptions. The data was analyzed to check the assumption for hierarchical linear modeling. For the final model, it was investigated whether the model assumptions have been met and also whether influential observations may be impacting model parameter estimates. Level-one residuals for the final model were assessed for multilevel model assumptions.

Research Question 3 Analysis

In order to examine question 3, four groups were created using a median split on laboratory measures of English WM and teacher ratings of WM. Teachers whose ratings matched with children’s actual score were considered accurate groups and those who did not match were considered inaccurate groups. A one-way MANOVA was run to examine whether groups differed on several achievement measures in English and Spanish.

CHAPTER IV – RESULTS

The results are separated into two sections: preliminary analyses and main analyses. Preliminary analyses include details regarding data preparation and cleaning, descriptive statistics, factor structure of WM measures and WMRS (teacher ratings of WM), creation of factor scores, the intercorrelations between teacher ratings of WM and factor scores, and testing of assumptions. The main analyses include the result as related to the three main research questions.

Preliminary Analyses

Data Preparation

To conduct the analyses for this study, the data were first entered into SAS statistical software for data screening. An examination of the data was conducted to assess for missing information and the results showed that missingness ranged from 1.5 percent to 4 percent, which was not a considerable missing data. Because the multilevel modeling analyses are robust to missingness (Snijders & Bosker, 2012), observations with missingness were kept in the data.

In examining the data for distribution, assumptions of univariate and multivariate normality were upheld for most variables based on the measures of skewness and kurtosis (Appendix A). Out of all the variables, seven variables (the memory measures) did not meet the normality assumption. In addition, a total of 65 univariate outliers were found within the data, which included scores that fell 3.5 standard deviations above or below the mean. Since the outliers consisted of 13 percent of the data, a winsorizing method was used to deal with outliers instead of deleting them. Extreme values above and below

3.5 standard deviations from the mean were replaced in the data with values corresponding to plus and minus 3.5 standard deviations as appropriate (Ghosh & Vogt, 2012). Most variables, apart from five, yielded skewness and kurtosis less than 3, and met the criteria for univariate normality. Five variables in the data set, English listening sentence span, Spanish conceptual span, Spanish listening sentence span, Spanish updating, and mapping and directions were transformed using a square root transformation to improve normality. Preliminary analyses were also conducted to ensure to no violation of assumptions of linearity and homoscedasticity.

In addition to data cleaning, raw scores on all continuous variables were converted to z scores to make all variable on the same scale. The rating scales on the WMRS and CTRS were also reverse coded for easier interpretation.

Descriptive Statistics

The means and standard deviations for all the variables are provided in Table 1. The descriptive statistics of the measures prior to winsorizing and transformation are presented in Appendix A. Additionally, the correlations among the manifest variables were checked for multicollinearity (Appendix B). The WMRS and Cognitive problem/ Inattention subscale of CTRS were highly correlated, *r* = .88, therefore, inattention (CTRS) was removed from the analysis.

Confirmatory Factor Analysis: WM

A confirmatory factor analysis was conducted to determine whether Baddeley’s multicomponent WM model (3-factor model) that included the phonological loop (STM), measured by forward digit span, word span, phonological span, backward digit span, executive system, measured by conceptual span, listening sentence span, updating, rhyming span, and visual-spatial sketchpad measured by matrix and mapping/direction tasks would best fit the data (Swanson et al., 2011). A three-factor WM model (executive, phonological loop, and visual-spatial sketchpad) was hypothesized to underlie the ten memory measures. A confirmatory factor analysis of the model provided a good fit to the data for the English 3-factor model, χ^2 (30) = 59.48, $p < .001$, $\chi^2/df = 1.98$, CFI = .95, non-normed = .92, RMSEA = .05. For the Spanish 3-factor model, the obtained goodness-of-fit indices were χ^2 (32) = 86.89, $p < .01$, $\chi^2/df = 2.72$, CFI = .90, non-normed = .85, RMSEA = .06 (lower confidence limit = .05), indicating an adequate model fit. The findings suggest that the three-factor model captures the same constructs within the two language systems.

After establishing the three-factor model as the better model for both language systems, a new model was tested that included both language memory measures together. A second order factor model was tested to account for the shared variance between the four language-specific memory factors (English STM, Spanish STM, English WM, and Spanish WM) along with Visual-Spatial WM factor (Swanson et al., 2004). However, the model was not a good fit for the data, χ^2 (129) = 378.37, $p < .001$, $\chi^2/df = 2.93$, CFI = .83, non-normed = .80, RMSEA = .06. Since the model with the second-order analysis did not have an adequate fit, two measurement models were selected, a separate three-factor model for English and Spanish.

Confirmatory Factor Analysis: WMRS

A confirmatory factor analysis was conducted to determine whether one factor underlie the 20 items on the WMRS. The goodness-of-fit indices indicated an adequate fit for one-factor model, χ^2 (170) = 1107.99, $p < .01$, $\chi^2/df = 6.52$, CFI = .91, non-normed= .90, RMSEA = .10. An RMSEA values between .08 and .10 considered a mediocre model fit (Hair, Anderson, Tatham, & Black, 1998). A composite variable was created for the WMRS and used as the variable for teacher ratings of WM.

Factor Scores

Due to large number of measures in the study, each set of measures with two or more variables were reduced to factor scores. Using the approach outlined by Nunnally & Bernstein (1994), factor scores were created by multiplying the z score of the target variable to the standardized factor loading weights based on the total sample, which was created using the PROC CALIS command in the SAS program. Means, standard deviations, and standardized factor loadings are reported in Appendix C. Factor scores were created for English and Spanish memory measures (executive WM, STM, and visual-spatial sketchpad), and English and Spanish reading and oral language.

Intercorrelations

The intercorrelations among the factor scores WM and achievement in both languages and WMRS were computed. Table 2 shows the intercorrelations among all the measures. Due to a large sample size, all correlations (*rs* > .15) except one were significant at an alpha of $p < .0001$. Instead, the focus was on the magnitude of correlations; coefficients of .50 were considered high, which are equivalent to Cohen's d of .80 (Cohen, 1988).

The primary predictor, teacher ratings of WM (WMRS), was moderately correlated with English reading, $r = .37$, and Spanish reading, $r = .36$. There were low correlations between WMRS and all achievement variables, ranging from $r = .15$ to $r = .24$. The magnitude of the correlations between teacher ratings and laboratory measures of WM were lower ($r = .18$ to $r = .24$) than the correlations between teacher ratings and reading ($r = .36$). WMRS was not significantly correlated with visual-spatial WM, $r = .05$, $p > .05$. When the effect of achievement was removed from the relationship between teacher ratings of WM and laboratory measures of WM, the magnitude of correlations decreased. The partial correlations between teacher ratings of WM and English Executive WM, English STM, and visual-spatial WM were $r = .06$, .06, and -.03, respectively when the effect of achievement measures was partialed out.

Correlations were also assessed for cross-language transfer, correlations between English and Spanish on the same constructs. Reading, oral language, executive WM, and short-term memory measures were assessed in both languages. High effect sizes ($rs > .50$) emerged on reading and STM. Moderate effect sizes ($rs < .50$ and $> .30$) emerged on oral language and executive WM.

There was a significant relationship between memory measures and achievement measures: Spanish executive WM was related to Spanish oral language ($r = .55$), English STM was related to English reading ($r = .54$) and English oral language ($r = .52$), and Spanish STM was related to Spanish oral language ($r = .50$). Moderate correlations were found between English executive WM and English reading and oral language skills, *rs* = .43 and .37, respectively. Low correlations were found between visual-spatial WM and most of the other variables, ranging from *rs* = .19 to .38.

Test of Assumptions

When conducting multilevel analyses, assumptions of residual normality, linearity, and homogeneity of variance must be upheld. To test the normality assumption, standardized residuals were plotted against their normal scores. The plot indicated that residuals fell in relatively straight line, indicating that the error terms are normally distributed. To assess assumptions of normality, linearity, and homogeneity of variance simultaneously, a plot of predicted values against the level-one residuals was examined. Roughly equivalent frequencies of points were scattered above and below the mean without making any particular shape such as funneling out; this provided evidence that the assumptions have not been violated (Raudenbush & Bryk, 2002). In addition, Levene’s test indicated equal variance of the level-one residuals across each level-two unit ($F = 1.09$, $p = .30$). Assumptions for the final model were examined only at level-one since there were no predictors at level-two.

Main Analyses

Research Questions 1 and 2

Question 1 - Do teacher ratings predict performance on laboratory measures of WM and do these predictions reflect the relationship between a general WM system or specific components of WM?

Question 2 - Are teacher ratings of WM confounded by student characteristics related to achievement, language proficiency, gender, and grade?

The criterion measures of this study were components of WM (phonological loop, central executive system, and visual-spatial) and a composite WM variable for both English and Spanish. The primary predictor variable was WMRS (teacher ratings of WM). In addition to the primary predictor of WMRS, the following covariates (student achievements measures of English and Spanish reading, oral language, and mathematics skills along with their English language proficiency [CELDT], gender, and grade) were entered into the multilevel regression model to determine if teacher ratings uniquely predict WM when partialed for the influence of these variables.

Results are presented for each dependent variable: English and Spanish overall WM (Table 3), English and Spanish executive WM (Table 4), English and Spanish STM (Table 5), and Visual-Spatial WM (Table 6).

English overall WM

Unconditional model. The left section of Table 3 presents the parameter estimates and standard errors for the three English WM models. The unconditional model (model 1a) estimated the intercept as (y_{00} = - 0.07, *se* = .15, *t* = -0.44), which is the average WM score across all classes and children. The variance of the child-level residual errors, was estimated as $\hat{\sigma}^2 = 4.03$, *se* = .28, *Z* = 14.61. The variance of the classroom-level residual errors was estimated as $\hat{\tau}_{00}$ = 0.98, *se* = .27, *Z* = 3.60. All parameter estimates for the random effects were larger than the corresponding standard errors, and calculation of the Z-test shows that they are all significant at *p* <.001. The intraclass correlation, calculated as 0.98/ (0.98 + 4.03) equaled 0.20. Thus, 20% of the variance in the English WM scores was at the classroom level. This indicated that the grouping according to classes led to an important similarity between the results of different students in the same class, although within class-differences was far larger than between-class differences. Since the unconditional model contains no explanatory variables, the residual variances represented unexplained error variance. The deviance reported in the table is a measure of model misfit, and when explanatory variables are added to the model, the deviance is expected to go down.

Conditional model (WMRS only). The conditional model (model 2a) in Table 3 included teacher ratings (WMRS) as an explanatory variable. The regression coefficient for WMRS was $y_{10} = 0.69$, *se* = .09, *t* = 7.55, which means that with each unit increase on the WMRS, the WM score is expected to increase by 0.69 units. The intercept for this model was $y_{00} = -0.09$, *se* = .16, *t* = -0.57, and not significant.

The deviance of the null model (Model 1a) was compared to the deviance of this model (Model 2a). The deviance of this model was 2041.4, which was a reduction of 2184.9-2041.4 = 135.5 from the deviance of the equivalent model without the teacher rating effect estimated (Model 1a). Models 1a and 2a differed by only a single parameter estimate ($\hat{\gamma}_{10}$), so the difference between these deviances was distributed as chi-square with a single degree of freedom: χ^2 (1) = 135.5, $p < .0001$.

In order to assess how much of the variance remaining in the children's WM scores unaccounted for by classroom differences was attributable to teacher ratings of WM, a proportion of variance reduction was calculated. The addition of teacher ratings of WM reduced the residual variance to 14 percent [(4.03-3.46)/4.03 = .14]. Fourteen percent of explainable variation in intercept value in WM was a function of WMRS.

In order to investigate whether the teacher ratings vary across classrooms, the random slope variance was also examined. WMRS was entered into the random effects along with the intercept. The variance of the regression coefficients for WMRS was estimated as 0. This variance component was not significant, so the hypothesis that the regression slopes for WMRS rating vary across classrooms was not supported by the data. To estimate whether slopes varied across classrooms, a likelihood ratio test comparing regression model with and without random slope variation was conducted. Random slopes showed no significant variation across classroom, [(2041.4 – 2041.4 = 0.01); χ^2 (2) = 0.1, $p > .05$]. Therefore, the residual variance term for the WMRS slopes was removed from the model as recommended by Snijders and Bosker (2012) to exclude nonsignificant random effects, and only the random intercept model were assessed.

Conditional model (all variables). The third model was an extension to Model 2a which included English reading, Spanish reading, English oral language, Spanish oral language, mathematics, English language proficiency (CELDT), gender, and grade as explanatory variables. When these variables were entered in the model, teacher ratings was no longer significant, indicating that teacher ratings do not uniquely predict WM ($y_{10} = 0.03$, $se = .10$, $t = 0.30$). There was a substantial shift in fixed effect intercept from Model 2a to Model 3a ($y_{00} = 0.05$, $se = 0.26$, $t = 0.19$), even though it remained non-significant. English reading ($y_{20} = 0.34$, $se = .09$, $t = 3.72$), English oral language ($y_{40} = 0.26$, $se = .07$, $t = 3.97$), and mathematics ($y_{60} = 0.41$, $se = .16$, $t = 2.59$) were positively and significantly related to children's English WM. However, there was no association between English WM and Spanish reading (y_{30} = 0.11, *se* = .07, *t* = 1.55), oral language (y_{50} = 0, *se* = .07, *t* = 0.02), CELDT (y_{70} = 0.29, *se* = .21, *t* = 1.35), gender (y_{80} = -0.21, *se* = .17, *t* = -1.29), second grade (y_{90} = 0.01, *se* = .27, *t* = 0.04), and third grade (y_{100} = -0.06, *se* = .39, *t* = -0.16), all *ps* > .05.

The residual variance for Model 3a was 2.76, *se* = .19, *Z* = 14.75. The proportion of variance calculation was (3.46 – 2.76)/ 3.46 = .20; this means that an addition of these explanatory variables to Model 2a reduced the within-classroom variance by 20%. When compared to the conditional model with only WMRS predictor, Model 3a deviance values indicated a significantly better fit to the data, $\Delta\chi^2(9)$ = 364.8 (2041.4 –1676.6), *p* < .001.

Spanish Overall WM

Unconditional model. The right section of Table 3 presents the parameter estimates and standard errors for the three Spanish overall WM models. The unconditional model (model 1b) estimated the intercept as (y_{00} = - 0.07, *se* = .16, *t* = -0.43), which was the average Spanish WM score across classrooms and children. The variance of the child-level residual errors was estimated as $\hat{\sigma}^2 =$ 4.74, *se* = .32, *Z* = 14.68. The variance of the classroom-level residual errors was estimated as $\hat{\tau}_{00}$ = 1.20, *se* = .32, *Z* = 3.78. The intraclass correlation, calculated as 1.20/ (1.20 + 4.74) equals 0.20. Accordingly, 20% of the variance in the Spanish WM scores was at the classroom level, which is same as the English overall WM.

Conditional model (WMRS only). The conditional model (model 2b) in Table 3 included teacher ratings (WMRS) as an explanatory variable. The regression coefficient for WMRS was $y_{10} = 0.78$, *se* = .10, *t* = 7.86, indicating that with each unit increase on the WMRS, the WM score was expected to increase by 0.78 units. The intercept for this model was $y_{00} = -0.08$, *se* = .18, *t* = -0.47, and not significant. The deviance of the null model (Model 1b) was compared to the deviance of this model (Model 2b). This model deviance values indicated a significantly better fit to the data, $\Delta\chi^2(1) = 135.8$ (2268 – 2132.2), $p < .0001$. Adding teacher ratings of WM reduced the residual variance to 14 percent [(4.74-4.09)/4.74 = .14]. Fourteen percent of explainable variation in intercept value in Spanish overall WM was a function of WMRS.

Conditional model (all variables). Similar to the English WM conditional model 3a, this model included all explanatory variables. When these variables were entered in the model, teacher ratings was no longer significant, indicating that teacher ratings do not uniquely predict WM ($y_{10} = 0.01$, *se* = .10, *t* = 0.11). Spanish reading ($y_{30} = 0.24$, *se* = .07, *t* = 3.18), English oral language ($y_{40} = 0.21$, *se* = .07, *t* = 3.08), Spanish oral language ($y_{50} = 0.43$, *se* = .07, *t* = 6.27), mathematics ($y_{60} = 0.48$, *se* = .17, *t* = 2.85), and gender ($y_{80} = -0.50$, *se* = .17, *t* = -2.90) were related to Spanish overall WM. Gender was a dummy coded variable (0 = female, 1 = male), which means that on average the boys scored 0.50 units lower than girls on the Spanish overall WM. There was no relationship found between Spanish overall WM and English reading ($y_{30} = 0.11$, *se* = .07, *t* = 1.74), CELDT ($y_{70} = 0.29$, *se* = .21, *t* = 1.17), second grade ($y_{90} = 0.01$, *se* = .27, *t* = -1.14), and third grade ($y_{100} = - 0.06$, *se* = .39, *t* = -1.06), all *ps* > .05. The residual variance for Model 3b was 2.92, *se* = .21, *Z* = 13.61. The proportion of variance calculation was (4.09 – 2.92)/ 4.09 = .29. This means that addition of these explanatory variables to Model 2b reduced the within-classroom variance by 29%. When compared to the conditional model with only WMRS predictor, Model 3b deviance values indicated a significantly better fit to the data, $\Delta\chi^2(9) = 319.9$ (2032.2 –1712.3), $p < .001$.

English Executive WM

Unconditional model. Table 4 (left section) presents the parameter estimates and standard errors for the three English executive WM models. The intercept for the unconditional model (model 1a) was (y_{00} = - 0.05, *se* = .05, *t* = -1.02), which is the average English executive WM score across classrooms and children. The variance of the child-level residual errors, was estimated as $\hat{\sigma}^2 = 0.48$, *se* = .03, *Z* = 14.66. The variance of the classroom-level residual errors was estimated as $\hat{\tau}_{00}$ = 0.10, *se* = .03, *Z* = 3.41. The intraclass correlation, calculated as 0.10/ (0.10 + 0.48) equals 0.17. Thus, 17% of the variance in the English executive WM scores was at the classroom level.

Conditional model (WMRS only). The conditional model (model 2a) in Table 4 included teacher ratings (WMRS) as an explanatory variable. The regression coefficient for WMRS was y_{10} = 0.17, *se* = .02, *t* = 5.17, which means that with each unit increase on the WMRS, the WM score was expected to increase by 0.17 units. The intercept for this model was y_{00} = -0.05, *se* = .05, *t* = -1.01, and not significant. The deviance of the null model (Model 1a) was compared to the deviance of this model (Model 2a). The deviance values for this model indicated a significantly better fit to the data, $\Delta\chi^2(1) = 66.6$ (1114.9 – 1048.3), $p < .001$. Adding teacher ratings of WM reduced the residual variance to 6 percent [(0.48-0.45)/0.48 = .06]. Six percent of explainable variation in intercept value in English executive WM was a function of WMRS.

Conditional model (all variables). Similar to the English overall WM conditional model 3a, this model included all explanatory variables. When these variables were entered in the model, teacher ratings was no longer significant, indicating that teacher ratings do not uniquely predict WM ($y_{10} = -0.01$, *se* = .04, *t* = -0.22). English reading ($y_{30} = 0.07$, *se* = .04, *t* = 1.97), mathematics ($y_{60} = 0.25$, *se* = .07, *t* = 3.83) were the only two variables predictive of English executive WM. All other variables were not related to English executive WM: Spanish reading ($y_{30} = 0.01$, *se* = .03, *t* = 0.40), English oral language ($y_{40} = 0.05$, *se* = .03, *t* = 1.92), Spanish oral language ($y_{50} = 0$, *se* = .03, *t* = 0.01), gender ($y_{80} = -0.04$, *se* = .07, *t* = -0.54), CELDT ($y_{70} = 0.10$, *se* = .09, *t* = 1.11), second grade ($y_{90} = -0.06$, *se* = .12, *t* = -0.49), and third grade ($y_{100} = -0.29$, *se* = .17, *t* = -1.71, all *ps* > .05. The residual variance for Model 3a was 0.42, *se* = .03, *Z* = 13.77. The proportion of variance calculation was (0.48 – 0.42)/ 0.48 = .13. This means that addition of these explanatory variables to Model 2a reduced the within-classroom variance by 13%. When compared to the conditional model with only WMRS predictor, Model 3a had significantly better fit to the data, $\Delta\chi^2(9) = 168.8$ (1048.3 –879.5), $p < .001$ than Model 2a.

Spanish Executive WM

Unconditional model. The right section of Table 4 presents the parameter estimates and standard errors for the three Spanish executive WM models. The intercept for the unconditional model (model 1b) was estimated as ($y_{00} = -\ 0.04$, *se* = .06, *t* = -

0.64), which was the average Spanish executive WM score across classrooms and children. The variance of the child-level residual errors was estimated as $\hat{\sigma}^2 = 0.85$, *se* = .06, *Z* = 14.55. The variance of the classroom-level residual errors was estimated as $\hat{\tau}_{00}$ = 0.16, *se* = .05, *Z*= 3.33. The intraclass correlation, calculated as 0.16/ (0.16 + 0.85) equals 0.16, which indicated that 16% of the variance in the Spanish executive WM scores was at the classroom level.

Conditional model (WMRS only). The conditional model (model 2b) in Table 4 included teacher ratings (WMRS) as an explanatory variable. The regression coefficient for WMRS was y_{10} = 0.27, *se* = .04, *t* = 6.26, which means that with each unit increase on the WMRS, the WM score is expected to increase by 0.27 units. The intercept for this model was y_{00} = -0.05, *se* = .06, *t* = -0.71, and not significant. The deviance of the null model (Model 1b) was compared to the deviance of this model (Model 2b). This model deviance values indicated a significantly better fit to the data, $\Delta\chi^2(1)$ = 86.8 (1370.9 – 1284.1), *p* < .0001. Adding teacher ratings of WM reduced the residual variance by nine percent [(0.85-0.77)/0.85 = .09]. Nine percent of explainable variation in intercept value in WM was a function of WMRS.

Conditional Model (all variables). Like other conditional models, model 3b included all explanatory variables. Similar to the findings of previous conditional models, teacher ratings was no longer significant, indicating that teacher ratings do not uniquely predict WM (y_{10} = -0.01, *se* = .05, *t* = -0.15). Similar to the results of Spanish overall WM, Spanish reading (y_{30} = 0.08, *se* = .03, *t* = 2.34), English oral language (y_{40} = 0.07, *se* = .073, *t* = 2.44), Spanish oral language (y_{50} = 0.20, *se* = .03, *t* = 6.50), and mathematics (y_{60} = 0.22, *se* = .08, *t* = 2.90) were related to Spanish executive WM. There was no relationship found between Spanish executive WM and English reading (y_{30} = 0.04, *se* = .04, *t* = 0.92), CELDT (y_{70} = 0.05, *se* = .10, *t* = 0.46), gender (y_{80} = -0.22, *se* = .08, *t* = -2.90), second grade (y_{90} = -0.17, *se* = .13, *t* = -1.29), and third grade (y_{100} = - 0.27, *se* = .19, *t* = -1.40), all *ps* > .05. The residual variance for Model 3b was 0.59, *se* = .04, *Z* = 13.72. The proportion of variance calculation was (0.77 – 0.59)/ 0.77 = .23. This means that addition of these explanatory variables to Model 2b reduced the within-classroom variance by 23%. When compared to the conditional model with only WMRS predictor, Model 3b deviance values indicated a significantly better fit to the data, $\Delta\chi^2(9)$ = 263.1 (1284.1 – 1021), *p* < .001.

English STM

Unconditional model. The left section of Table 5 presents the parameter estimates and standard errors for the three English STM models. The unconditional model (model 1a) estimated the intercept as (y_{00} = 0, *se* = .10, *t* = -0.01), which is the average English STM score across classrooms and children. The variance of the child-level residual errors was estimated as $\hat{\sigma}^2 =$ 1.96, *se* = .13, *Z* = 14.58, and the variance of the classroom-level residual errors was estimated as $\hat{\tau}_{00}$ = 0.37, *se* = .11, *Z* = 3.34. The intraclass correlation, calculated as 0.16 [0.37/ (0.37 + 1.96)]. Sixteen percent of the variance in the English STM scores was at the classroom level.

Conditional model (WMRS only). The conditional model (model 1a) in Table 5 included teacher ratings (WMRS) as an explanatory variable. The regression coefficient for WMRS was y_{10} = 0.47, *se* = .06, *t* = 7.34, which meant that as the ratings on the

WMRS increased by one unit, the English STM score was expected to increase by 0.47 units. The intercept for this model was $y_{00} = -0.02$, *se* = .10, *t* = -0.20, and not significant. The deviance of the null model (Model 1a) was compared to the deviance of this model (Model 2a). This model deviance values indicated a significantly better fit to the data, $\Delta\chi^2(1) = 128.4$ (1785.8 – 1657.4), $p < .001$. Adding teacher ratings of WM reduced the residual variance to 14 percent [(1.96-1.68)/1.96 = .14]. Fourteen percent of explainable variation in the intercept value in WM was a function of WMRS.

Conditional model (all variables). Similar to the English overall WM conditional model 3a, this model included all explanatory variables. When these variables were entered in the model, teacher ratings was no longer significant, indicating that teacher ratings do not uniquely predict STM ($y_{10} = 0.12$, *se* = .08, *t* = 1.58). The only variables that predicted English STM were English reading ($y_{30} = 0.20$, *se* = .07, *t* = 2.93) and English oral language ($y_{40} = 0.16$, *se* = .05, *t* = 3.24).

The Spanish achievement measures, mathematics, and student characteristics did not predict English STM: Spanish reading ($y_{30} = 0.08$, *se* = .05, *t* = 1.54), Spanish oral language ($y_{50} = -0.02$, *se* = .05, *t* = -0.35), mathematics ($y_{60} = 0.09$, *se* = .12, *t* = 0.73), CELDT ($y_{70} = 0.14$, *se* = .16, *t* = 0.91), gender ($y_{80} = -0.19$, *se* = .12, *t* = -1.55), second grade ($y_{90} = 0.15$, *se* = .21, *t* = 0.72), and third grade ($y_{100} = 0.29$, *se* = .30, *t* = 0.98, all *ps* > .05.

The residual variance for Model 3a was 1.46, *se* = .11, *Z* = 13.95. The proportion of variance calculation was (1.68 – 1.46)/ 1.68 = .13. This means that addition of these explanatory variables to Model 2a reduced the within-classroom variance by 13%. When compared to the conditional model with only WMRS predictor, Model 3a deviance values indicated a significantly better fit to the data, $\Delta\chi^2(9) = 246.2$ (1657.4 – 1411.2), $p < .001$.

Spanish STM

Unconditional model. The right section of Table 5 presents the parameter estimates and standard errors for the three Spanish STM models. The intercept for the unconditional model (model 1b) was ($y_{00} = - 0.01$, *se* = .09, *t* = -0.08), which was the average Spanish STM score across classrooms and children. The variance of the child-level residual errors was estimated as $\hat{\sigma}^2 = 1.92$, *se* = .13, *Z* = 14.71. The variance of the classroom-level residual errors was estimated as $\hat{\tau}_{00} = 0.31$, *se* = .10, *Z* = 3.15. The intraclass correlation, calculated as 0.31/ (0.31 + 1.92) equals 0.14, indicating that 14% of the variance in the Spanish STM scores was at the classroom level.

Conditional model (WMRS only). The conditional model (model 2b) in Table 5 includes teacher ratings (WMRS) as an explanatory variable. The regression coefficient for WMRS was $y_{10} = 0.44$, *se* = .06, *t* = 6.97, indicating that with each unit increase on the WMRS, the WM score was expected to increase by 0.44 units. The intercept for this model was $y_{00} = -0.01$, *se* = .10, *t* = -0.08, and not significant. The deviance of the null model (Model 1b) was compared to the deviance of this model (Model 2b). This model deviance values indicated a significantly better fit to the data, $\Delta\chi^2(1) = 114.1$ (1796.9 – 1682.8), $p < .0001$. Adding teacher ratings of WM reduced the residual variance to 11 percent [(1.92-1.70)/1.92 = .11]. Eleven percent of explainable variation in the intercept value in Spanish STM was a function of WMRS.

Conditional model (all variables). Similar to the English STM conditional model 3a, this model included all explanatory variables. When these variables were entered in the model, teacher ratings was no longer significant, indicating that teacher ratings do not uniquely predict WM (y_{10} = 0.09, *se* = .08, *t* = 1.25). Similar to the results of Spanish overall WM, Spanish reading (y_{30} = 0.14, *se* = .05, *t* = 2.67), Spanish oral language (y_{50} = 0.21, *se* = .05, *t* = 4.41), and gender (y_{80} = -0.29, *se* = .12, *t* = -2.42) were related to Spanish STM. This means that on average the boys scored 0.29 units lower than girls on the STM measures. However, unlike Spanish WM model, English oral language (y_{40} = 0.07, *se* = .05, *t* = 1.60), and mathematics (y_{60} = 0.20, *se* = .12, *t* = 1.70) were not related to Spanish STM. There was also no relationship found between Spanish STM and English reading (y_{30} = 0.07, *se* = .07, *t* = 1.09), CELDT (y_{70} = 0.16, *se* = .15, *t* = 1.06), second grade (y_{90} = -0.12, *se* = .21, *t* = -0.56), and third grade (y_{100} = - 0.17, *se* = .30, *t* = -0.56), all *ps* > .05.

The residual variance for Model 3b was 1.34, *se* = .10, *Z* = 13.46. The proportion of variance calculation was (1.70 – 1.34)/ 1.70 = .21. This means that the addition of these explanatory variables to Model 2b reduced the within-classroom variance by 21%. When compared to the conditional model with only WMRS predictor, Model 3b deviance values indicated a significantly better fit to the data, $\Delta\chi^2(9)$ = 294.2 (1682.8 – 1388.61), *p* < .001.

Visual-Spatial WM

Unconditional model. Table 6 presents the parameter estimates and standard errors for the Visual-Spatial WM models. The unconditional model (model 1) estimated the intercept as (y_{00} = - 0.01, *se* = .04, *t* = -0.28), which is the average visual-spatial WM score across classrooms and children. The variance of the child-level residual errors was estimated as $\hat{\sigma}^2$ = 0.39, *se* = .03, *Z* = 14.65. The variance of the classroom-level residual errors was estimated as $\hat{\tau}_{00}$ = 0.04, *se* = .02, *Z* = 2.64. The intraclass correlation, calculated as 0.04/ (0.04 + 0.39) equals 0.09. Thus, nine percent of the variance in the visual-spatial WM scores was at the classroom level.

Conditional model (WMRS only). Only teacher ratings (WMRS) were entered in this model (model 2) as an explanatory variable. The regression coefficient for WMRS was y_{10} = 0.05, *se* = .03, *t* = 1.60. Unlike previous WM and STM models, WMRS was not a significant predictor. The intercept for this model was y_{00} = -0.01, *se* = .04, *t* = -0.29, and not significant. The deviance of the null model (Model 1) was compared to the deviance of this model (Model 2). This model deviance values indicated a significantly better fit to the data, $\Delta\chi^2(1)$ = 39.1 (967.4 – 928.3), *p* < .001. However, adding teacher ratings of WM did not reduce the residual variance as it was same as unconditional model, σ^2 = 0.39.

Conditional model (all variables). Even though teacher ratings of WM were not significant in Model 2, other explanatory variables were still added to the model to examine whether any of those predicted visual-spatial WM. Only one variable was significant predictor of visual-spatial WM, English oral language (y_{40} = 0.05, *se* = .02, *t* = 2.28). English reading (y_{30} = 0.05, *se* = .03, *t* = 1.63), Spanish reading (y_{30} = 0.02, *se* = .02, *t* = 0.91), Spanish oral language (y_{50} = 0.02, *se* = .02, *t* = 0.85), mathematics (y_{60} = 0.06, *se* = .06, *t* = 0.99), CELDT (y_{70} = 0.04, *se* = .08, *t* = 0.58), gender (y_{80} = 0.03, *se* =

.06, t = 0.42), second grade (y_{90} = -0.14, *se* = .10, t = -0.14), and third grade (y_{100} = -0.02, *se* = .14, t = 0.17) were not related to visual-spatial WM, all *ps* > .05.

The residual variance for Model 3 was 0.36, *se* = .02, Z = 14.73. The proportion of variance calculation was (0.39 - 0.36)/ 0.39 = .08. This means that addition of these explanatory variables to Model 2 reduced the within-classroom variance by 8 percent. When compared to the conditional model with only WMRS predictor, Model 3 deviance values indicated a significantly better fit to the data, $\Delta\chi^2(9)$ = 145.4 (928.3 – 782.9), p < .001.

Research Question 3

Question 3: Do teachers who accurately identify WM problems differ from teachers who incorrectly identify WM problems on student measures of English and Spanish reading, English and Spanish oral language, mathematics achievement, and age?

Four groups were created based on the median scores for performance on the English overall WM variable (laboratory measures) and overall teacher ratings on the WMRS. A median split was used on the laboratory measure of WM and teachers ratings of WM because Katz (2006) recommended that a median split is a good choice where there is no natural cut-off. Groups were categorized as 0, 1, 2, and 3. The hypothesis of interest was whether the groups differed on measures of student achievement and language skills. It was predicted that groups less accurate in their predictions were more likely to be influenced by student performance on measures of achievement and language than the accurate group. That is, groups 1 and 2 were more likely to be influenced by the extremes in student achievement than the accurate group.

Group 0 included children who scored low on WM and had lower teacher ratings; whereas, Group 3 included children who had high performance on WM measures and also received high ratings. Groups 0 (low-low) and 3 (high –high) were considered high accuracy groups because teachers' ratings matched with children's performance on WM measures. In contrast, groups 1 and 2 were considered low accuracy groups in which teachers ratings did not match children's performance on laboratory WM measures. Group 1 (low-high) included children who had low performance on WM but received high ratings from teachers, and Group 2 (high-low) consisted of children who received low ratings from teachers even though they had higher scores on WM measures. The creation of four groups allowed to examine whether teachers who accurately identified WM problems in children (Group 0) differed from those who were unable to identify WM problems (Groups 1 and 2). The sizes of the groups were relatively equivalent. Groups 0, 1, 2, and 3 had 136, 107, 102, and 137 students, respectively. More students were in the accurate groups (groups 0 and 3) than non-accurate groups.

A one-way multivariate analysis of variance (MANOVA) was conducted to determine whether there is a statistical and practical differences among accuracy groups on multiple dependent variables (English and Spanish reading, English and Spanish oral language, mathematics, and age). A MANOVA was selected over multiple ANOVAs because it regards the linear combination of dependent variables, and takes into account the interrelationship between dependent variables.

Significant differences were found among the four groups on the joint distribution of the dependent variables, Wilks' $\lambda = .52$, F (18, 1338.3) = 19.17, $p < .0001$. Table 7 presents the means and standard deviations of the dependent variables for the four groups. Univariate analyses of variance (ANOVAs) for each dependent variable were conducted as follow-up tests to the MANOVA. Using the Bonferroni method for controlling Type I error rates for multiple comparisons, each ANOVA was tested at the .05/6 = .008 level. Post hoc analyses for the dependent variables were also conducted given the statistically significant ANOVA *F* tests. Specifically, Tukey HSD tests were conducted on all possible pairwise contrasts. Group 0 (accurate group) was compared to Group 1 and 2 (non-accurate groups). Also, Group 3 (accurate group) was compared to non-accuracy groups, Groups 1 and 2. Since four comparisons were made for each dependent variable, each pairwise comparison was tested at the .008/4, or .0002, significance level.

In general, the results suggest that the accurate group were more influenced by the extremes in student performance than the inaccurate group, suggesting that the match between teacher ratings and WM performance was mediated by extremes in student achievement and language. As evident in Table 7, the means on all variables are much lower for children in Group 0 (accurate group) and much higher for children in Group 3 (accurate group); whereas, the means were in middle range for children in inaccurate groups (1 and 2).

English reading. The ANOVA of the English reading was significant, F (3,478) = 79.59, p <.0001, $\eta^2 = .33$. Tukey's HSD comparisons indicated a significant difference between all groups ($p < .0001$): group 0 (low WM-low rating; M = -1.28, SD = 1.58), group 1 (high WM - low rating; M = 0.48, SD = 1.28), group 2 (low WM - high rating; M

= -0.51, *SD* = 1.58), and group 3 (high WM – high rating; *M* = 1.28, *SD* = 1.31). In other words, groups that were considered accurate (groups 0 and 3), meaning teachers' ratings of WM matched to students' actual WM scores were separable from groups that were inaccurate on English reading. Group 0 produced significantly poor performance on the English reading measures in comparison to other groups.

Spanish reading. The ANOVA based on the Spanish reading was also significant, F (3,478) = 43.72, p <.0001, η^2 = .22. Tukey's HSD comparisons indicated a significant difference between all groups (p < .0001): group 0 (*M* = -0.87, *SD* = 1.0), group 1 (*M* = -0.12, *SD* = 1.41), group 2 (*M* = -0.16, *SD* = 1.64), and group 3 (*M* = 1.13, *SD* = 1.73). Similar to English reading variable, children in Group 0 performed poorly on the Spanish reading measures compared to other three groups.

English oral language. Similarly, the ANOVA of the English oral language was significant, F (3,478) = 62.20, p <.0001, η^2 = .28. The following pairs of groups were found to be significantly different (p < .002): groups 0 (*M* = -1.21, *SD* = 1.91) and 1 (*M* = 0.83, *SD* = 1.66), and groups 3 (*M* = 1.31, *SD* = 1.79) and 2 (*M* = -0.98, *SD* = 1.81). However, no significant difference were found between groups 0 and 2 (p = .30), and groups 3 and 1 (p = .04). The results indicated that Group 0 (accurate group) was only different from Group 1 but not from Group 2. This indicates that English oral language performance of children who have low WM and low ratings is similar to those who have low WM and high ratings.

Spanish oral language. The ANOVA was statistically significant for Spanish oral language, F (3,478) = 14.49, p <.0001, η^2 = .08. Similar to the findings for English oral language, the following pairs of groups were found to be significantly different (p < .002): groups 0 (M = -0.55, SD = 1.56) and 1 (M = 0.24, SD = 1.65), and groups 3 (M = 0.62, SD = 1.58) and 2 (M = -0.29, SD = 1.53). However, no significant difference were found between groups 0 and 2 (p = .20), and groups 3 and 1 (p = .10). The results were similar to English oral language, in that, Group 0 (accurate group) was only different from Group 1 but not from Group 2. Group 0 and Group 2 (non-accurate group) performed similarly on Spanish oral language.

Mathematics. The ANOVA based on the mathematics was also significant, F (3,478) = 41.10, p <.0001, η^2 = .21. Groups 0 (M = -0.49, SD = 0.82) and 1 (M = 0.42, SD = 0.87) were different from each other (p < .002) and groups 3 (M = 1.58, SD = 0.47) and 2 (M = -0.40, SD = 0.87) were different from each other (p < .002). However, no significant difference were found between groups 0 and 2 (p = .50), and groups 3 and 1 (p = .72). In other words, children in Group 0 had lower performance on mathematics compared to Group 1, but similar performance to Group 2.

Age. In addition, children's age was also added in to model as a dependent variable to examine whether groups differ on children's age. The ANOVA for this model was significant, F (3,478) = 34.30, p <.0001, η^2 = .18. The following pairs of groups were found to be significantly different (p < .002): groups 0 (Mean Age in months = 88.63, SD = 10.83) and 1 (M = 97.09, SD = 8.55), and groups 3 (M = 94.31, SD = 10.05) and 2 (M = 85.07, SD = 8.90). However, no significant difference were found between groups 0 and 2 (p = .005), and groups 3 and 1 (p = .03). The findings suggested that age was not related to the group differences because while mean age for Group 0 (accurate group) was relatively lower than Group 1 (non-accurate group), it was similar to Group 2 (non-accurate group).

Research Question 3: Follow-up Analysis

For research question 3, a follow-up analysis was conducted in which variables related to student achievement and characteristics were entered into a multilevel regression model to predict teachers' WM ratings.

Unconditional model. The left section of Table 8 presents the parameter estimates and standard errors for the unconditional model. The unconditional model estimated the intercept as (y_{00} = 0.01, *se* = .05, *t* = 0.22), which is the average teacher ratings of WM across classrooms and children. The variance of the child-level residual errors was estimated as $\hat{\sigma}^2 = 0.94$, *se* = .07, *Z* = 14.32, and the variance of the classroom-level residual errors was estimated as $\hat{\tau}_{00}$ = 0.05, *se* = .04, *Z* = 1.54. The intraclass correlation, calculated as 0.05 [0.05/ (0.05 + 0.94)]. Five percent of the variance in the teacher WM ratings was at the classroom level.

Conditional model. The conditional model in Table 8 included student achievement and characteristics variables. The conditional model estimated the intercept as (y_{00} = 1.11, *se* = .13, *t* = 8.45), which is the average teacher ratings of WM across classrooms and children. Other than English and Spanish oral language, all variables related to student achievement and characteristics were predictive of teachers' ratings of WM. English reading (y_{30} = 0.27, *se* = .04, *t* = 6.85), Spanish reading (y_{30} = 0.09, *se* = .03, *t* = 2.70), mathematics (y_{60} = 0.36, *se* = .07, *t* = 4.99), CELDT (y_{70} = 0.29, *se* = .09, *t* = 3.12), gender (y_{80} = -0.24, *se* = .07, *t* = -3.41), second grade (y_{90} = -1.10, *se* = .15, *t* =

-7.20), and third grade (y_{100} = -1.83, *se* = .19, *t* = -9.54) were related to teachers' ratings of WM., all *ps* < .05. Only English oral language (y_{40} = 0.01, *se* = .03, *t* = 0.41) and Spanish oral language (y_{50} = 0.01, *se* = .03, *t* = 0.46) were not related to teacher WM ratings.

The residual variance for the conditional model was 0.48, *se* = .04, *Z* = 13.46. The proportion of variance calculation was (0.94 - 0.48)/ 0.94 = .49. This means that addition of these explanatory variables to the unconditional model reduced the within-classroom variance by 49 percent. When compared to the unconditional model, the deviance values of the conditional model indicated a significantly better fit to the data, $\Delta\chi^2(9)$ = 372.2 (1354.6 – 982.4), *p* < .001.

CHAPTER V – DISCUSSION

The purpose of this study was to identify whether teacher ratings of EL children's WM is predictive of laboratory measures of WM and whether this relationship still holds after controlling for student achievement and characteristics. In addition, the study explored whether teacher ratings differed based on the reading, language, and mathematics skills of Spanish-speaking ELs. The results of the study are summarized addressing the major research questions of the study.

Question 1: Do teacher ratings predict performance on laboratory measures of WM and do these predictions reflect the relationship between a general WM system or specific components of WM?

The results suggest that teacher ratings of WM do predict the performance on laboratory measures of WM. This relationship was evident for both the general WM system and specific components of WM in both language systems, but not for the visual-spatial WM. While, the study supports the findings of previous studies that the WMRS, a classroom observation scale for teachers, is an accurate indicator of WM deficits in children (Alloway et al., 2009b; Normand & Tannock, 2014); these findings may only generalize to WM tasks that incorporate verbal skills. This study qualifies previous studies by suggesting that the WMRS was not predictive of visual-spatial WM. No doubt, my findings depend on the way WM constructs was operationalized. In this study, visual-spatial WM construct was measured by two tasks: visual-matrix and mapping & directions. Whereas, in Alloway et al. (2009b), the laboratory measures of WM included the Automated Working Memory Assessment (AWMA) and the Working Memory Index from the WISC-IV. In their study, three measures of visual-spatial WM included an “odd-one-out task”, the “Mr X task”, and the “spatial recall task”. While the nature of the variables is similar to visual-spatial measures used this study, it is possible that the results of this study are influenced due to the measures used in as well as how they were scored.

The results from this study extend the literature in three ways. First, the relationship between teacher ratings and WM is explored for children whose second language is English. Previous studies (e.g., Alloway et al., 2009b; Normand & Tannock, 2014) that have explored the relationship between teacher ratings and WM have only studied this relationship with monolingual children. Previous studies by Alloway et al. (2009b) included children from England, and Normand and Tannock (2014) study had children from Canada whose primary language was English. Thus, the contrasts in finding may be related to the students’ English language proficiency.

Second, this study varies from the findings of Alloway and colleagues in that it explored the relationship of teacher ratings to an overall general WM construct as well as different components of WM. Separation analysis for the WM components allowed to determine that even though teacher ratings are predictive of overall WM, when partialed out, it is only related to executive WM and STM for both English and Spanish.

Finally, the findings determine whether the relationship between teacher ratings on an observable scale and laboratory measures of WM holds when children’s achievement measures are included in the model. The findings suggest that the relationship between teacher observation and laboratory measures do not hold when achievement measures are included in the analysis.

Question 2: Are teacher ratings of WM confounded by student characteristics related to achievement, language proficiency, gender, and grade?

As indicated previously, teacher ratings of WM were confounded by student achievement and students characteristics. For all WM models, the addition of the following covariates (student achievements measures of English and Spanish reading, oral language, and mathematics skills along with their English language proficiency [CELDT], gender, and grade) eliminated the significance of teacher ratings.

The results reveal that teacher ratings on EL children is not necessarily based on their observations on children's WM abilities, but rather, their ratings might be influenced by the observable behaviors in the classroom related to inattentive behavior or academic performance of reading, oral language, and mathematics skills (Guzman- Orth et al, 2014). The relationship between WM and inattention is evident in many studies; however, Alloway et al. (2009b) found that teacher ratings on the WMRS identified more children with WM problems than the teacher ratings on the CTRS, a classroom behavior scale with inattention subscale. While, Alloway and colleagues found that WM (as measured by WMRS) is related to some aspects of attention, the findings of this study suggest that children's inattentive behavior might be influencing teacher ratings of WM.

As suggested by Alloway et al. (2009b), children's temperament and motivation could also influence teachers' ratings on the WMRS. It is possible that EL children who are shy or unmotivated might be identified by the teachers as having WM deficits based on the WMRS. For example, a child could "abandon activities before completion"

(WMRS item 4) or “not able to focus during activities” (WMRS item 18) due to lack of motivation rather than a child’s lower WM ability.

It is also possible that teacher ratings could be influenced by the number of EL children in the classroom. In the study, there was a large variation in the number of ratings scales teachers had to complete, ranging from 1 to 20. It is also possible that a teacher who rated only one child might have been able to draw closer attention to that child’s classroom behavior than a teacher rating 20 children. Given the number of students per class, it would be impractical to expect teachers to have perfectly accurate insights into the WM capacity of each of their students.

The findings from question two show that not only did the achievement variables eliminated the significance of teacher ratings in predicting WM, but different variables were significant in different models. Though the purpose of the study was not to determine the variables that are related to WM, but interesting findings occurred for different models. The only variables to predict English WM were English reading, English oral language, and mathematics; whereas, for Spanish WM, there was a cross-language transfer. Along with mathematics, Spanish reading and oral language, English language was also related to Spanish WM. Interestingly, gender was also related to Spanish WM, findings suggest that boys underperformed on Spanish WM than girls.

The results for English executive WM and English STM also differed from the overall English WM. English oral language was not related to English executive WM, but, it was related to English short-term memory; whereas, mathematics was related to executive WM but not STM. The findings for the Spanish executive WM are similar to overall Spanish WM, in which Spanish reading, both Spanish and English oral language, and mathematics were related to Spanish executive WM. For Spanish STM, there were no cross-language transfers as only Spanish reading, Spanish oral language, and gender were significant predictors. The study's findings are similar to Swanson et al. (2011)'s in that weak cross-language transfer was found, and English literacy measures best predict English WM and Spanish measures best predict Spanish WM. However, in their study, WM measures were used to predict literacy; whereas, in this study laboratory measures of WM serve as dependent variables. Unlike the rest of the models, only English oral language was related to visual-spatial sketchpad.

The fact that different achievement measures predicted different components of WM could be due to different achievement measures tapping either storage (STM) or both storage and processing components (executive WM). Oral language consisted of vocabulary and syntax measures that were more related to recall of items than processing information; whereas, reading variables included both word identification and comprehension that involves both recall and processing of information (Swanson et al., 2004; Swanson et al., 2011).

Question 3: Do teachers who accurately identify WM problems differ from teachers who incorrectly identify WM problems on student measures of English and Spanish reading, English and Spanish oral language, mathematics achievement, and age?

In order to investigate this question, four groups were created to determine if teachers who are accurate in their identification of children's WM differ from teachers who are inaccurate in their identification on student measures of achievement. If teacher ratings matched children's WM scores, they were considered accurate in their identification, and if the teachers' ratings did not match children's WM scores, then they were considered inaccurate. These four groups were compared on measures of English and Spanish reading, oral language, mathematics and age. The results indicated that groups were statistically separable on measures of English and Spanish reading. However, results also indicated that groups were more comparable on English and Spanish oral language, mathematics, and age. These findings indicate that high accuracy groups were only separable from one inaccuracy group but not both. For example, on English oral language, Group 0 (children who had low WM and received low teacher ratings) was statistically different from Group 1 (children with high WM but low teacher ratings), but not from Group 2 (children with low WM but high ratings). Similar pattern occurred for Spanish oral language, mathematics, and age.

These findings also show that the classroom behaviors listed on the WMRS are not necessarily accurate indictors of WM deficits. One hundred and two children were rated by their teachers as having working memory deficits, however they had average scores on the laboratory WM measures. In addition, 107 children received higher ratings on the WMRS, when they had lower WM scores on the laboratory measures of WM. The findings suggest that teacher ratings were influenced by student performance on achievement and language measures because children in accurate groups (groups 0 and 3) had low and high performance on the achievement measures, respectively.

The findings of this study leads to a new question as to whether the ratings of cognition (WM) are in fact those of cognition or are they influenced by other factors such as achievement. Are teacher WM ratings a surrogate for achievement? The findings for this study challenge whether classroom observation measures are capturing a cognitive construct (e.g., WM, inattention) as opposed to individual differences in achievement. For example, Fuchs et al. (2006) studied the association of cognitive variables (e.g., working memory, processing speed, phonological decoding, attentive behavior, long-term memory, language ability, and reading skills) with mathematical performance. The interesting finding of the study was that inattention (teacher ratings of children's attentive behavior) uniquely predicted all three mathematical skills, while partialling out the influence of the cognitive variables. The authors speculated that this could be due to teacher ratings function "as a proxy for achievement rather than index attention and/or distractibility" (p. 38).

Similar to Fuchs et al., the findings of this study suggest that teacher ratings on the WMRS could be a surrogate for achievement rather than WM. Gathercole et al. (2006) noted that "working memory deficits are not easy to detect on the basis of informal contact alone and may easily be misclassified either as attentional problems or more pervasive cognitive impairments" (p. 234). WM is not an observable entity, and the items on the rating scale are a translation of behaviors that are associated with WM deficits. Therefore, it is possible that due to the strong relationship between WM deficits and lower achievement, teachers might be rating students based on their classroom performance and behavior rather than children's actual WM ability. The findings are supported by Pimperton and Nation's study (2014), which found that children who were classified as poor comprehenders (i.e., children who have age-appropriate reading skills but impairments in reading comprehension) were rated by their teachers to show significantly higher frequency of problem behaviors on the WMRS. Similarly, in this study, the results suggest that teachers rated children as having WM deficits based on their scores on achievement measures rather than WM ability.

It is also unclear from this study whether the teacher ratings of WM provide additional information above the objective laboratory measures of WM. Even though teacher ratings might be a proxy for academic achievement, it does offer teachers assistance in identifying children struggling with learning.

Limitations

Several limitations need to be kept in mind when interpreting the results of this study. First, variables related to teacher characteristics are not included in the multilevel models due to large amount of missing data. Future research is warranted to evaluate factors that may influence teachers' judgment accuracy on WMRS, including teacher characteristics (e. g., teaching experience, gender, and familiarity with WM). Second, only Spanish-speaking EL children are included in the study; therefore, findings cannot be generalized to ELs who first language is other than Spanish or English monolingual children. Third, in order to analyze the third research question, accuracy groups were created using an arbitrary cutoff at the median split on English overall WM and teacher ratings. Fourth, factor scores were used in this study, and it is recommended that future studies should use latent variables to take into account the measurement error.

Fifth, the factor structure of the WMRS was determined using a confirmatory factor analysis approach, and a lenient fit criteria (RMSEA > .08) was used. The fit indices indicate that the model should be modified and reevaluated for better fit; however, since the focus of the study was not to explore the factor structure of WMRS, only one factor was retained, and a total score of the items was created to use as the primary predictor. It is recommended that the items on the WMRS should be explored further to determine whether certain items are more or less correlated with laboratory measures of WM. Lastly, due to the co-occurrence of inattention in children with WM deficits (Alloway et al., 2009b), it was proposed that teachers' ratings on students' classroom inattention behavior would also be included in the models; however, in this study, teachers' ratings on the inattention subscale of CTRS was highly correlated with teacher ratings of WM, and it was decided to not include in the analyses.

Implications

As Alloway and colleagues (2009b) and this study found WM problems are difficult to detect from classroom observations alone. Gathercole and Alloway (2008) recommend that teachers should evaluate the WM demands of classroom activities and be able to identify particular features in an activity that places larger demands on WM; thus, help identify EL children with WM problems. Working memory teacher rating scale may be useful tool for early identification of EL children with achievement and/or WM problems; however, it is not a stand-alone tool, and does not replace laboratory measures of WM. It does offer teachers assistance for recognizing classroom behaviors that are linked to poor WM, and thus help identify EL children with WM problems.

Summary and Future Research

Taken together, these findings suggest that teacher ratings are a proxy for achievement rather than ratings of cognition. Future research that includes the development of observational measure that predicts low verbal and visual-spatial WM in order to bypass the language issues may help teachers better identify children with WM deficits. Although WMRS includes some items related to visual-spatial WM such as “loses his or her place in complicated activities” (WMRS item 12) and “mixes up material inappropriately, e.g., incorrectly combines parts from two sentences rather than reading each one accurately” (WMRS item 6); future research might benefit from inclusion of items that separates verbal WM and visual-spatial WM.

References

Alloway, T. P. (2006). How does working memory work in the classroom? *Educational Research and Reviews, 1*(4), 134-139.

Alloway, T. P. (2010). *Improving working memory: Supporting students' learning*. California: SAGE Publications Ltd.

Alloway, T. P., Doherty-Sneddon, G., & Forbes, L. (2012). Teachers' perceptions of classroom behavior and working memory. *Education Research and Review, 7*(6), 138-142.

Alloway, T. P., Gathercole, S. E., Adams, A.-M., Willis, C., Eaglen, R., & Lamont, E. (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. *British Journal of Developmental Psychology, 23*, 417-426.

Alloway, T. P., Gathercole, S. E., & Kirkwood, H. J. (2008). *Working memory rating scale*. London: Pearson Assessment.

Alloway, T. P., Gathercole, S. E., Kirkwood, H. J., & Elliot, J. (2009a). The cognitive and behavioral characteristics of children with low working memory. *Child Development*, *80*(2), 606-621. doi:10.111/j.1467-8624.2009.01282.x

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliot, J. (2009b). The working memory rating scale: A classroom-based behavioral assessment of working memory. *Learning and Individual Differences, 19,* 242-245. doi:10.1016/j.lindif.2008.10.003

Alloway, T. P., Gathercole, J. H., Holmes, J., Place, M., Elliot, J. G., & Hilton, K. (2009). The diagnostic utility of behavioral checklists in identifying children with ADHD and children with working memory deficits. *Child Psychiatry Human Development, 40,* 353-366.

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in early childhood. *Journal of Experimental Child Psychology, 87,* 85-106.

Alloway, T. P., Pickering, S. J., & Gathercole, S. E. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*(6), 1698-1716.

Artiles, A. J., & Ortiz, A. A. (Eds.). (2002). *English language learners with special education needs*. Washington, DC: Center for Applied Linguistics and Delta Systems.

Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., Tahan, K. (2011). *The Condition of Education 2011* (NCES 2011-033). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Aud, S., Wilkinson-Flicker, S., Kristapovich, P., Rathbun, A., Wang, X., and Zhang, J. (2013). *The Condition of Education 2013* (NCES 2013-037). U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved May 20, 2014 from http://nces.ed.gov/pubsearch.

August. D., & Hakuta, K. (Eds.) (1997). *Improving schooling for minority-language children: A research agenda.* Washington, DC: National Academy Press.

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the national literacy panel on language minority children and youth*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Baddeley, A. (2006). Working memory: An overview. In S. Pickering (Ed.) *Working memory and education* (pp.1-31). San Diego, CA: Elsevier Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.). *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York: Cambridge University Press.

Berg, D. H. (2008). Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading. *Journal of Experimental Child Psychology*, *99*(1), 288-308. doi: 10.1016/j.jecp.2007.12.002

Betts, J. R., Rueben, K. S., & Danenberg, A. (2000). *Equal resources, equal outcomes? The distribution of school resources and student achievement in California.* San Francisco: Public Policy Institute of California.

Brownell, K. (2001). *Expressive One-Word Picture Vocabulary Test* (3rd Edition). New York: Academic Therapy Publications.

Cain, K. (2006). Children's reading comprehension: The role of working memory in normal and impaired development. In S. Pickering (Ed.) *Working memory and education* (pp.61-91). San Diego, CA: Elsevier Press.

Cain, K., Oakhill, J. V. & Bryant, P. E. (2000). Phonological skills and comprehension failure: A test of the phonological processing deficit hypothesis. *Reading and Writing, 13,* 31-56.

California Department of Education (2015). California English language development test-CalEdFacts [Fact Sheet]. Retrieved from http://www.cde.ca.gov/ta/tg/el/cefceldt.asp

Chiappe, P., Siegel, L. S., & Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading, 6*, 369-400.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2^{nd} ed.). New York: Academic Press.

Conners, C. K. (1997). *Conners' Rating Scales—Revised technical manual.* North Tonawanda, NY: Multi-Health Systems.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450-466. doi:10.1016/S0022-5371(80)90312-6

Dehn, M. J. (2008). *Working memory and academic learning: Assessment and intervention*. New Jersey: Wiley & Sons, Inc.

DeStefano, D., & LeFevre, Jo-A. (2010). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*(3), 353-386. doi: 10.1080/09541440244000328

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

Dunn, L. M., Lugo, D. E., Padilla, E. R., & Dunn, L. M. (1986). *Test de Vocabulario Imágenes Peabody.* Circle Pines, MN: American Guidance Service.

Eckert, T. L., Dunn, E. K., Codding, R. S., Begeny, J. C., Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the School, 43*(3), 247-265. doi: 10.1002/pits.20147

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research, 84*(5), 253-262.

Engel de Abreu, P. M. J., & Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology, 104*(4), 974-986. doi: 10.1037/a0028390

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., Schatschneider, C., & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*, 29-43. doi:10.1037/0022-0663.98.1.29

Fry, R. (2007). *How far behind in math and reading are English language learners?* Washington, DC: Pew Hispanic Center.

Gathercole, S. E., & Alloway, T. P. (2008). *Working memory and learning: A practical guide to teachers*. Thousand Oaks, CA: SAGE Publications Ltd.

Gathercole, S. E., Alloway, T. P., Kirkwood, H. J., Elliot, J. G., Holmes, J., & Hilton, K. A. (2008). Attentional and executive function behaviors in children with poor working memory. *Learning and Individual Differences, 18*, 214-223.

Gathercole, S. E., Alloway, T. P., Willis, C., Adams, A. M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology, 93*, 265-281.

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory abilities and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology*, 22, 1019-1037.

Gathercole, S. E., Lamont, E., & Alloway, T. P. (2006). Working memory in the classroom. In S. Pickering (Ed.) *Working memory and education* (pp.219-241). San Diego, CA: Elsevier Press.

Gathercole, S. E. & Pickering, S. J. (2000). Assessment of working memory in six- and seven- year- old children. *Journal of Educational Psychology, 92*(2), 377-390.

Gathercole, S. E. & Pickering, S. J. (2001). Working memory deficits in children with special educational needs. *British Journal of Special Education, 28*, 89-97.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177-190.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from National Curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology, 40*, 1-16.

Geva, E., & Massey-Garrison, A. (2013). A comparison of the language skills of ELLs and monolinguals who are poor decoders, poor comprehenders, or normal readers. *Journal of Learning Disabilities, 46*(5), 387-401. doi: 10.1177/0022219412466651

Geva, E., & Siegel, L. S. (2000). Orthographic and cognitive factors in the concurrent development of basic reading skills in two languages. *Reading and Writing: An Interdisciplinary Journal, 12*, 1-30.

Ghosh, D., & Vogt, A. (2012). Outliers: An evaluation of methodologies. *Joint Statistical Meetings*, 3455-3460.

Gottardo, A., Stanovich, K. E., & Siegel, L. S. (1996). The relationships between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third-grade children. *Journal of Experimental Child Psychology, 63*, 563-582.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6-10. doi:10.1177/074193258600700104

Guzman-Orth, D., Grimm, R., Gerber, M., Orosco, M., Swanson, H. L., & Lussier, C. (2014). Psychometric properties of the working memory rating scale for Spanish-speaking English language learners. *Journal of Psychoeducational Assessment,* 1-13. doi: 10.1177/0734282914558710

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate Data Analysis* (5th ed.). New Jersey: Prentice-Hall.

Hammill, D., Mather, N., & Roberts, R. (2001). *Illinois Test of Psycholinguistic Abilities, Third Edition.* Austin, TX: PRO-ED.

Hemphill, F. C., & Vanneman, A. (2011). *Achievement Gaps: How Hispanic and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress* (NCES 2011-459). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Jerman, O., Reynolds, C., & Swanson, H. L. (2012). Does growth in working memory span or executive processes predict growth in reading and math in children with reading disabilities? *Learning Disability Quarterly, 35*(3), 144-157. doi: 10.1177/0731948712444276

Jongejan, W., Verhoeven, L., & Siegel, L. S. (2007). Predictors of reading and spelling abilities in first- and second-language learners. *Journal of Educational Psychology, 99*(4), 835-851. doi: 10.1037/0022-0663.99.4.835

Kail, R., & Hall, L. K. (2001). Distinguishing short-term memory from working memory. *Memory and Cognition, 29* (1), 1–9. doi:10.3758/BF03195735.

Katz, M. H. (2006). *Study design and statistical analysis: A practical guide for clinicians*. New York: Cambridge University Press.

Kofler, M.J., Rapport, M.D., Bolden, J., Sarver, D.E., & Raiker, J.S. (2010). ADHD and working memory: The impact of central executive deficits and exceeding storage/rehearsal capacity on observed inattentive behavior. *Journal of Abnormal Child Psychology*, 38, 149-161.

Lesaux, N.K., Crosson, A., Kieffer, M.J., & Pierce, M. (2010). Uneven profiles: Language minority learners' word reading, vocabulary, and reading comprehension skills. *Journal of Applied Developmental Psychology, 31,* 475-483.

Lesaux, N.K., Lipka, O., & Siegel, L.S. (2006). Investigating cognitive and linguistic abilities that Influence the reading comprehension skills of children from diverse linguistic backgrounds. *Reading & Writing: An Interdisciplinary Journal, 19,* 99-131

Lesaux, N. K., & Siegel, L. S. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology, 39*(6), 1005-1019. doi: 10.1037/0012-1649.39.6.1005.

Limbos, M., & Geva, E. (2001). Accuracy of teacher assessments of second language students at risk for reading disability. *Journal of Learning Disabilities, 34*(2), 136-151.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2013). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic*

Bulletin & Review, 1-23. doi:10.3758/s13423-013-0565-2

Mass, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86-92. doi: 10.1027/1614-1881.1.3.86

McCardle, P., Mele-McCarthy, J., Cutting, L., Leos, K., & D'Emilio, T. (2005). Learning disabilities in English language learners: Identifying the issues. *Learning Disabilities Research & Practice*, *20*, 1 – 5.

Mih, V., & Mih, C. (2011). The role of working memory deficits in children with poor comprehension ability. *Procedia – Social and Behavioral Sciences, 29*, 247-355. doi: 10.1016/j.sbspro.2011.11.249

Normand, S., & Tannock, R. (2014). Screening for working memory deficits in the classroom: The psychometric properties of the working memory rating scale in a longitudinal school-based study. *Journal of Attention Disorders, 18*(4), 294-304.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Pickering, S. J., & Gathercole, S. E. (2004). Distinctive working memory profiles in children with special education needs. *Journal of Educational Psychology, 24*, 392-408.

Pimperton, H., & Nation, K. (2014). Poor comprehenders in the classroom: Teacher ratings of behavior in children with poor reading comprehension and its relationship with individual differences in working memory. *Journal of Learning Disabilities, 47*(3), 199-207.

Rapport, M.D., Bolden, J., Kofler, M.J., Sarver, D.E., Raiker, J.S., Alderson, R.M. (2009). Hyperactivity in boys with attention-deficit/hyperactivity disorder (ADHD): A ubiquitous core symptom or manifestation of working memory deficits? *Journal of Abnormal Child Psychology, 37,* 521-534.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Application and data analysis methods, second edition*. Thousand Oaks, CA: Sage.

Raykov, T., & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. Mahwah, NJ: Lawrence Erlbaum.

Salvesen, K. A., & Undheim, J. O. (1994). Screening for learning disabilities with teacher rating scales. *Journal of Learning Disabilities, 27*(1), 60-61.

Samson, J. F., & Lesaux, N. K. (2009). Language-minority learners in special education: Rates and predictors. *Journal of Learning Disabilities, 42*(2), 148- 162. doi: 10.1177/0022219408326221

SAS Institute. (2015). The SAS system for Windows. Release 9.4. Cary, NC.

Saunders, W. M., & Marcelleti, D. J. (2013). The gap that can’t go away: The catch-22 of reclassification in monitoring the progress of English learners. *Education Evaluation and Policy Analysis, 35*, 139-157. doi: 10.3102/0162373712461849.

Sideridis, G. D., Antonious, F., & Padeliadu, S. (2008). Teacher biases in the identification of learning disabilities: An application of the logistic multilevel model. *Learning Disabilities Quarterly, 31*(4), 199-209.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development, 60,* 973-980.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Los Angeles: SAGE Publications Ltd.

Speece, D. L., Ritchey, K. D., Silverman, R., Schatshneider, C., Walker, C. Y., & Andrusik, K. N. (2010). Identifying children in middle childhood who are at risk for reading problems. *School Psychology Review, 39*(2), 258-276.

St. Clair-Thompson, H. L. (2011). Executive functions and working memory behaviors in children with a poor working memory. *Learning and Individual Differences, 21,* 409-414.

Steele, M. M. (2004). Making the case for early identification and intervention for young children at risk for learning disabilities. *Early Childhood Education Journal, 32*(2), 75-79.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology. Advance online publication.* doi: 10.1037/a0027627

Swanson, H. L. (1992). Generality and modifiability of working memory among skilled and less skilled readers. *Journal of Educational Psychology, 84*, 473-488. doi:10.1037//0022-0663.84.4.473

Swanson, H. L. (1996a). Individual and age-related differences in children’s working memory. *Memory & Cognition, 24, 70*-82. doi:10.3758/BF03197273

Swanson, H. L. (1996b). *S-Cognitive Processing Test (S-CPT): A Dynamic Assessment Measure*. Austin, TX: PRO-ED.

Swanson, H. L. (2008). Working memory and intelligence in children: What develops? *Journal of Educational Psychology, 100(3)*, 581-602. doi:10.1037/0022-0663.100.3.581

Swanson, H.L. (2011). *Abbreviated Test of Working Memory: Experimental Edition*. Riverside, CA: University of California-Riverside, GSOE, Working Memory Research Center (WM+)

Swanson, H. L., Ashbaker, M. H., & Lee, C. (1996). Learning-disabled readers' working memory as a function of processing demands. *Journal of Experimental Child Psychology, 61*, 242-275. doi:10.1006/jecp.1996.0016

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not a risk for serious math difficulties. *Journal of Educational Psychology, 96,* 471-491. doi:10.1037/0022-0663.96.3.471

Swanson, H. L., & Berninger, V. (1995). The role of working memory in skilled and less skilled readers' comprehension. *Intelligence, 21,* 83-108. doi:10.1016/0160-2896(95)90040-3

Swanson, H. L., Howard, C. B., & Sáez, L. (2006). Do different components of working memory underlie different subgroups of reading disability? *Journal of Learning Disabilities, 39*, 252-269. doi: 10.1177/00222194060390030501

Swanson, H. L., & Howell, M. (2001). Working memory, short-term memory, and speech rate as predictors of children’s reading performance at different ages. *Journal of Educational Psychology*, *93,* 720-734.

Swanson, H.L., Mink, J., & Bocian, K.M. (1999). Cognitive processing deficits in poor readers with symptoms of reading disabilities and ADHD: More alike than different? *Journal of Educational Psychology, 91*, 321-333.

Swanson, H. L., Orosco, M., & Lussier, C. M. (2011). Cognition and literacy in English language learners at risk for reading disabilities. *Journal of Educational Psychology, 104*(2), 302-320. doi:10.1037/a0026225

Swanson, H. L., Orosco, M., & Lussier, C. M. (2015). Growth in literacy, cognition, and working memory in English language learners. *Journal of Experimental Child Psychology, 132*(1), 155-188.

Swanson, H. L., Orosco, M. J., Lussier, C. M., Gerber, M. M., & Guzman-Orth, D. A. (2011). The influence of working memory and phonological processing on English language learner children's bilingual reading and language acquisition. *Journal of Educational Psychology, 103*(4), 838-856. doi:10.1037/a0024578

Swanson, H. L., Sáez, L., & Gerber, M. (2006). Growth in literacy and cognition in bilingual children at risk or not at risk for reading disabilities. *Journal of Educational Psychology, 98*(2), 247-264. doi:10.1037/0022-0663.98.2.247

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),* 3-18. doi:10.1037/0022-0663.96.1.3

Swanson, H. L., & Zheng, X. (2013). Memory difficulties in children and adults with learning disabilities. In H. L. Swanson, K. Harris, & S. Graham. *Handbook of learning disabilities* (2nd ed.). New York, NY: Guilford Press.

Taylor, H., Anselmo, M., Foreman, A.L., Schatschneider, C., & Angelopoulos, J. (2000). Utility of kindergarten teacher judgments in identifying early learning problems. *Journal of Learning Disabilities, 33*, 200–210

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, *National Assessment of Educational Progress (NAEP), various years*, *1990–2013 Mathematics and Reading Assessments*. Retrieved from http://www.nationsreportcard.gov/

Wagner, R. K., Francis, D. J., & Morris, R. D. (2005). Identifying English language learners with learning disabilities: Key challenges and possible approaches. *Learning Disabilities Research and Practice, 20*(1), 17–23

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-III*. San Antonio, TX: Psychological Corporation.

Wilkinson, G. S. (1993). *WRAT3 Administration Manual.* Delaware: Wide Range.

Woodcock, R. W., Muñoz-Sandoval, A. F. & Alverado, C. G. (2005). *Woodcock-Muñoz Language Survey*. Itasca, IL: Riverside Publishing.

Zheng, X., Swanson, H. L., & Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of Experimental Child Psychology, 110*(4), 481–498. doi: http://dx.doi.org/10.1016/j.jecp.2011.06.001

Table 1

Descriptive Statistics for all Measures after Winsorizing and Transformation

Variables (raw scores)	N	Mean	SD	Skewness	Kurtosis
Chronological Age (in months)	500	91.21	10.67	---	---
Fluid intelligence – Raven	462	22.60	6.50	-0.28	-0.10
Teacher Ratings					
Teachers' WM Ratings (WMRS)	479	43.08	16.38	-0.74	-0.66
CTRS (Inattention)	479	10.18	4.73	-0.67	-0.86
Reading English					
Letter Word Identification	487	34.98	11.35	0.06	-0.72
Comprehension	489	12.86	5.12	-0.52	-0.62
Reading Spanish					
Letter Word Identification	481	25.15	12.5	1.33	2.76
Comprehension	486	7.02	4.03	0.89	-0.07
Oral Language English					
Syntax	483	8.51	5.89	0.25	-1.15
Expressive Vocabulary	488	47.79	13.55	-0.39	1.16
Receptive Vocabulary	487	97.46	10.03	-0.06	0.08
Oral Language Spanish					
Syntax	485	7.93	6.69	0.24	-1.52
Expressive Vocabulary	488	28.20	16.71	-0.27	-1.13
Receptive Vocabulary	485	45.06	16.52	-0.50	1.19
Mathematics					
Raw score	487	22.98	4.49	0.44	-0.65
Short-Term Memory - English					
Forward Digit Span Task	489	3.55	0.91	-0.23	1.18
Backward Digit Span Task	487	2.08	0.95	-0.71	0.82
Phonetic Memory Span Task	492	1.20	0.69	0.26	0.05
Real-Word Span Task	491	2.32	0.78	-0.19	0.25
Short-Term Memory - Spanish					
Forward Digit Span Task	488	3.35	0.82	-0.33	2.09
Backward Digit Span Task	487	1.76	0.97	-0.72	-0.01
Phonetic Memory Span Task	500	1.18	0.77	-0.32	1.70
Real-Word Span Task	491	1.78	0.84	0.04	-0.36
Executive WM – English					
Conceptual Span Task	500	3.08	3.08	0.71	2.50
Listening Sentence Span	490	1.42	0.41	0.74	0.03
Rhyming Word Span	500	3.30	3.26	1.33	1.10
Updating	492	2.19	1.90	0.77	1.74

Executive WM - Spanish					
Conceptual Span Task	483	1.63	0.48	0.49	0.37
Listening Sentence Span	490	1.26	0.30	0.67	-0.76
Rhyming Word Span	483	2.96	2.53	1.15	1.11
Updating	491	1.47	0.58	0.72	-0.66
Visual-Spatial WM					
Visual matrix	490	11.14	6.69	0.23	0.16
Mapping/direction	490	1.40	0.55	1.47	2.22

Note. Only raw scores are reported.

Table 2

Intercorrelations between WMRS, Achievement, and Working Memory

	Variable	1	2	3	4	5	6	7	8	9	10	11
1	Teachers' WM Ratings (WMRS)	---										
2	E-Reading	.37	---									
3	S-Reading	.36	.61	---								
4	E-Oral Lang	.22	.75	.41	---							
5	S-Oral Lang	.15	.38	.58	.35	---						
6	Mathematics	.16	.68	.51	.59	.45	---					
7	E-Executive WM	.18	.43	.28	.37	.16	.39	---				
8	S-Executive WM	.21	.47	.49	.43	.55	.48	.37	---			
9	E-STM	.24	.54	.38	.52	.27	.45	.42	.44	---		
10	S- STM	.22	.42	.47	.37	.50	.40	.27	.54	.58	---	
11	Visual-Spatial	$.05^{ns}$	.38	.27	.36	.22	.36	.19	.26	.26	.20	---

Note. 1 = Teacher ratings of WM (WMRS); 2 = English reading; 3 = Spanish reading; 4 = English oral language; 5 = Spanish oral language; 6 = Mathematics; 7 = English executive WM; 8 = Spanish executive WM; 9 = English STM; 10 = Spanish STM; 11 = visual-spatial WM. All $p < .001$ unless indicated; ns = $p > .05$

Table 3

Multilevel Model Predicting Children's English and Spanish Overall WM

	English Overall WM			Spanish Overall WM		
	Model 1a	Model 2a	Model 3a	Model 1b	Model 2b	Model 3b
Variable	β (SE)	β (SE)	β (SE)	β (SE)	β (SE)	β (SE)
Fixed Effects						
Intercept, $\hat{\gamma}_{00}$	-0.07 (.15)	-0.09 (.16)	0.05 (.26)	-0.07 (.16)	-0.08 (.18)	0.51 (.28)
Teachers' WM ratings, $\hat{\gamma}_{10}$		0.69*** (.09)	0.03 (.10)		0.78*** (.10)	0.01 (.10)
E- Reading, $\hat{\gamma}_{20}$			0.34*** (.09)			0.17 (.10)
S- Reading, $\hat{\gamma}_{30}$			0.11 (.07)			0.24** (.07)
E- Oral Language, $\hat{\gamma}_{40}$			0.26*** (.07)			0.21** (.07)
S- Oral Language, $\hat{\gamma}_{50}$			0.00 (.07)			0.43*** (.07)
Mathematics, $\hat{\gamma}_{60}$			0.41** (.16)			0.48** (.17)
CELDT _high, $\hat{\gamma}_{70}$			0.29 (.21)			0.26 (.22)
Gender: male, $\hat{\gamma}_{80}$			-0.21 (.17)			-0.50** (.17)
Grade_2nd, $\hat{\gamma}_{90}$			0.01 (.27)			-0.33 (.29)
Grade_3rd, $\hat{\gamma}_{100}$			-0.06 (.39)			-0.45 (.42)
Random Effects	Variance (SE)			Variance (SE)		
Classroom, $\hat{\tau}_{00}$	0.98*** (.27)	1.32*** (.33)	0	1.20 *** (.32)	1.62*** (.39)	0.09 (.10)
Level-1 Residual, $\hat{\sigma}^2$	4.03*** (.28)	3.46*** (.24)	2.76*** (.19)	4.74*** (.32)	4.09*** (.29)	2.92*** (.21)
Model Fit Statistics						
Deviance (-2LL)	2184.9	2041.4	1676.6	2268.0	2132.2	1712.3
AIC	2190.9	2049.4	1700.6	2274.0	2131.2	1738.3
BIC	2198.0	2058.8	1728.6	2281.1	2140.6	1768.6

Note. Model 1a: ICC = 0.24. Model 1b: ICC = 0.20. **p* < .05. ***p* < .01. ****p* < .001.

Table 4

Multilevel Model Predicting Children's English and Spanish Executive WM

	English Executive WM			Spanish Executive WM		
	Model 1a	Model 2a	Model 3a	Model 1b	Model 2b	Model 3b
Variable	β (SE)	β (SE)	β (SE)	β (SE)	β (SE)	β (SE)
Fixed Effects						
Intercept, $\hat{\gamma}_{00}$	-0.05 (.05)	-0.05 (.05)	0.09 (.11)	-0.04 (.06)	-0.05 (.07)	0.24 (.13)
Teachers' WM ratings, $\hat{\gamma}_{10}$		0.17*** (.02)	-0.01 (.04)		0.27*** (.04)	-0.01 (.05)
E- Reading, $\hat{\gamma}_{20}$			0.07* (.04)			0.04 (.04)
S- Reading, $\hat{\gamma}_{30}$			0.01 (.03)			0.08* (.03)
E- Oral Language, $\hat{\gamma}_{40}$			0.05 (.03)			0.07* (.03)
S- Oral Language, $\hat{\gamma}_{50}$			0.00 (.03)			0.20*** (.03)
Mathematics, $\hat{\gamma}_{60}$			0.25*** (.07)			0.22** (.08)
CELDT _high, $\hat{\gamma}_{70}$			0.10 (.09)			0.05 (.10)
Gender: male, $\hat{\gamma}_{80}$			-0.04 (.07)			-0.22 (.08)
Grade_2nd, $\hat{\gamma}_{90}$			-0.06 (.12)			-0.17 (.13)
Grade_3rd, $\hat{\gamma}_{100}$			-0.29 (.17)			-0.27 (.19)
Random Effects	Variance (SE)			Variance (SE)		
Classroom, $\hat{\tau}_{00}$	0.10*** (.03)	0.12*** (.03)	0.03* (.02)	0.16 *** (.05)	0.21*** (.06)	0.02 (.02)
Level-1 Residual, $\hat{\sigma}^2$	0.48*** (.03)	0.45*** (.03)	0.42*** (.03)	0.85*** (.06)	0.77*** (.05)	0.59*** (.04)
Model Fit Statistics						
Deviance (-2LL)	1114.9	1048.3	879.5	1370.9	1284.1	1021.0
AIC	1120.9	1056.3	905.5	1376.9	1292.1	1047.0
BIC	1128.0	1065.7	935.8	1384.0	1301.5	1077.3

Note. Model 1a: ICC = 0.17. Model 1b: ICC = 0.16. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 5

Multilevel Model Predicting Children's English and Spanish STM

	English STM			Spanish STM		
	Model 1a	Model 2a	Model 3a	Model 1b	Model 2b	Model 3b
Variable	β (SE)	β (SE)	β (SE)	β (SE)	β (SE)	β (SE)
Fixed Effects						
Intercept, $\hat{\gamma}_{00}$	-0.00 (.10)	-0.02 (.10)	-0.09 (.20)	-0.01 (.09)	-0.01 (.10)	0.24 (.20)
Teachers' WM ratings, $\hat{\gamma}_{10}$		0.47*** (.06)	0.12 (.08)		0.44*** (.06)	0.09 (.08)
E- Reading, $\hat{\gamma}_{20}$			0.20** (.07)			0.07 (.06)
S- Reading, $\hat{\gamma}_{30}$			0.08 (.05)			0.14** (.05)
E- Oral Language, $\hat{\gamma}_{40}$			0.16*** (.05)			0.07 (.05)
S- Oral Language, $\hat{\gamma}_{50}$			-0.02 (.05)			0.21*** (.05)
Mathematics, $\hat{\gamma}_{60}$			0.09 (.12)			0.20 (.12)
CELDT _high, $\hat{\gamma}_{70}$			0.14 (.16)			0.16 (.15)
Gender: male, $\hat{\gamma}_{80}$			-0.19 (.12)			-0.29** (.12)
Grade_2nd, $\hat{\gamma}_{90}$			0.15 (.20)			-0.12 (.21)
Grade_3rd, $\hat{\gamma}_{100}$			0.29 (.30)			-0.17 (.30)
Random Effects	Variance (SE)			Variance (SE)		
Classroom, $\hat{\tau}_{00}$	0.37*** (.11)	0.48*** (.13)	0.04 (.04)	0.31 *** (.10)	0.45*** (.12)	0.10* (.02)
Level-1 Residual, $\hat{\sigma}^2$	1.96*** (.13)	1.68*** (.12)	1.46*** (.11)	1.92*** (.13)	1.70*** (.12)	1.34*** (.10)
Model Fit Statistics						
Deviance (-2LL)	1785.8	1657.4	1411.2	1796.9	1682.8	1388.6
AIC	1791.8	1665.4	1437.2	1802.9	1690.8	1414.6
BIC	1798.9	1674.8	1467.5	1810.0	1700.2	1444.9

Note. Model 1a: ICC =.16. Model 1b: ICC = 0.14. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 6

Multilevel Model Predicting Children's Visual-Spatial WM

	Visual-Spatial WM		
	Model 1	Model 2	Model 3
Variable	β (SE)	β (SE)	β (SE)
Fixed Effects			
Intercept, $\hat{\gamma}_{00}$	-0.01 (.04)	-0.01 (.04)	-0.03 (.09)
Teachers' WM ratings, $\hat{\gamma}_{10}$		0.05 (.03)	-0.06 (.04)
E- Reading, $\hat{\gamma}_{20}$			0.05 (.03)
S- Reading, $\hat{\gamma}_{30}$			0.02 (.03)
E- Oral Language, $\hat{\gamma}_{40}$			0.05* (.02)
S- Oral Language, $\hat{\gamma}_{50}$			0.02 (.02)
Mathematics, $\hat{\gamma}_{60}$			0.06 (.06)
CELDT _high, $\hat{\gamma}_{70}$			0.04 (.08)
Gender: male, $\hat{\gamma}_{80}$			0.03 (.06)
Grade_2nd, $\hat{\gamma}_{90}$			-0.01 (.10)
Grade_3rd, $\hat{\gamma}_{100}$			0.02 (0.14)
Random Effects	Variance (SE)		
Classroom, $\hat{\tau}_{00}$	0.04** (.02)	0.05** (.02)	0
Level-1 Residual, $\hat{\sigma}^2$	0.39*** (.03)	0.39*** (.03)	.36***(.02)
Model Fit Statistics			
Deviance (-2LL)	967.4	928.3	782.9
AIC	973.4	936.3	806.9
BIC	980.5	945.7	834.9

Note. Model 1: ICC = .09. **p* < .05. ***p* < .01. ****p* < .001.

Table 7

One-Way MANOVA Results

	Low WM/ Low Rating (n = 136)	Low WM/ High Rating (n = 107)	High WM/ Low Rating (n = 102)	High WM/ High Rating (n = 137)		
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*F*	η^2
E-Reading	-1.28 (1.58)	0.48 (1.28)	-0.51 (1.58)	1.28 (1.31)	79.59***	.33
S-Reading	-0.87 (1.00)	-0.12 (1.41)	-0.16 (1.64)	1.13 (1.73)	43.72***	.22
E- Oral Lang.	-1.21 (1.91)	0.83 (1.66)	-0.98 (1.81)	1.31 (1.79)	62.20***	.28
S- Oral Lang.	-0.55 (1.56)	0.24 (1.65)	-0.29 (1.53)	0.62 (1.58)	14.49***	.08
Mathematics	-0.49 (0.82)	0.42 (0.87)	-0.40 (0.87)	1.58 (0.47)	41.10***	.21
Age	88.63 (10.83)	97.09 (8.55)	85.07 (8.90)	94.31 (10.05)	34.30***	.18

Note. df = (3, 478). ***p < .0001. E = English, S = Spanish.

Table 8

Multilevel Model Predicting Teachers' Ratings of WM

	Unconditional Model	Conditional Model
Variable	β (SE)	β (SE)
Fixed Effects		
Intercept, $\hat{\gamma}_{00}$	0.01 (.05)	1.11*** (.13)
E- Reading, $\hat{\gamma}_{20}$		0.27*** (.04)
S- Reading, $\hat{\gamma}_{30}$		0.09** (.03)
E- Oral Language, $\hat{\gamma}_{40}$		0.01 (.03)
S- Oral Language, $\hat{\gamma}_{50}$		0.01 (.03)
Mathematics, $\hat{\gamma}_{60}$		0.36*** (.07)
CELDT _high, $\hat{\gamma}_{70}$		0.29** (.09)
Gender: male, $\hat{\gamma}_{80}$		-0.24*** (.07)
Grade_2nd, $\hat{\gamma}_{90}$		-1.10*** (.15)
Grade_3rd, $\hat{\gamma}_{100}$		-1.83*** (.19)
Random Effects	Variance (SE)	
Classroom, $\hat{\tau}_{00}$	0.05 (.04)	0.14***(.04)
Level-1 Residual, $\hat{\sigma}^2$	0.94*** (.07)	0.48***(.04)
Model Fit Statistics		
Deviance (-2LL)	1354.6	982.4
AIC	1360.6	1006.4
BIC	1367.7	1034.7

Note. ICC for unconditional model = .05. *p < .05. **p < .01. ***p < .001.

Appendix A

Descriptive Statistics for all Measures

Variables (raw scores)	N	Mean	SD	Skewness	Kurtosis
Chronological Age (in months)	500	91.21	10.67	---	---
Fluid intelligence – Raven	462	22.60	6.50	-0.28	-0.10
Teacher Ratings					
Teachers' WM Ratings (WMRS)	479	43.08	16.38	-0.74	-0.66
CTRS (Inattention)	479	10.18	4.73	-0.67	-0.86
Reading English					
Letter Word Identification	487	34.98	11.35	0.06	-0.72
Comprehension	489	12.86	5.12	-0.52	-0.62
Reading Spanish					
Letter Word Identification	481	25.15	12.50	1.33	2.76
Comprehension	486	7.02	4.03	0.89	-0.07
Oral Language English					
Syntax	483	8.51	5.89	0.25	-1.15
Expressive Vocabulary	488	47.79	13.55	-0.39	1.16
Receptive Vocabulary	487	97.46	10.03	-0.06	0.08
Oral Language Spanish					
Syntax	485	7.93	6.69	0.24	-1.52
Expressive Vocabulary	488	28.20	16.71	-0.27	-1.13
Receptive Vocabulary	485	45.06	16.52	-0.50	1.19
Mathematics					
Raw score	487	22.98	4.49	0.44	-0.65
Short-Term Memory - English					
Forward Digit Span Task	489	3.55	0.91	-0.23	1.18
Backward Digit Span Task	487	2.08	0.95	-0.71	0.82
Phonetic Memory Span Task	492	1.20	0.69	0.26	0.05
Real-Word Span Task	491	2.32	0.78	-0.19	0.25
Short-Term Memory - Spanish					
Forward Digit Span Task	488	3.35	0.82	-0.33	2.09
Backward Digit Span Task	487	1.76	0.97	-0.72	-0.01
Phonetic Memory Span Task	491	1.24	0.77	2.28	20.76
Real-Word Span Task	491	1.78	0.84	0.04	-0.36
Executive WM – English					
Conceptual Span Task	491	3.27	2.86	1.73	3.12
Listening Sentence Span	490	1.21	1.38	1.71	4.16
Rhyming Word Span	500	3.30	3.26	1.33	1.10
Updating	492	2.19	1.90	0.77	1.74

Executive WM - Spanish					
Conceptual Span Task	483	1.91	1.85	2.32	10.47
Listening Sentence Span	490	0.69	0.85	1.27	2.54
Rhyming Word Span	483	2.96	2.53	1.15	1.11
Updating	491	1.49	2.00	1.39	2.29
Visual-Spatial WM					
Visual matrix	490	11.14	6.69	0.23	0.16
Mapping/direction	490	1.36	2.63	4.80	33.44

Note. Descriptive data before winsorizing the variables and transformation.

Appendix B

Intercorrelations Between all Manifest Variables

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Teachers’ WM Ratings	--															
2	CTRS_Inattention	.88	--														
3	E-Letter Word ID	.38	.46	--													
4	E-Comprehension	.33	.41	.84	--												
5	S-Letter Word ID	.33	.36	.64	.48	--											
6	S-Comprehension	.35	.39	.60	.46	.77	--										
7	E-Syntax	.23	.30	.73	.70	.44	.45	--									
8	E-Expressive Vocab	.18	.22	.52	.58	.27	.26	.56	--								
9	E-Receptive Vocab	.16	.19	.61	.64	.32	.28	.70	.66	--							
10	S-Syntax	.16	.17	.33	.36	.35	.41	.38	.27	.29	--						
11	S-Expressive Vocab	.14	.12	.29	.27	.47	.53	.28	.19	.18	.49	--					
12	S-Receptive Vocab	.07	.10	.29	.26	.43	.43	.30	.19	.24	.36	.57	--				
13	Mathematics	.16	.19	.68	.63	.48	.48	.62	.41	.52	.34	.37	.39	--			
14	E-Conceptual Span	.22	.21	.41	.40	.33	.33	.39	.28	.38	.24	.22	.19	.38	--		
15	E-Listening Sentence	.11	.15	.29	.31	.20	.22	.36	.23	.28	.17	.22	.26	.32	.29	--	
16	E-Updating	.17	.19	.41	.35	.24	.23	.38	.21	.32	.14	.12	.14	.36	.33	.23	--
17	E-Rhyming	.12	.16	.29	.25	.17	.19	.28	.10	.19	.03	.08	.08	.25	.28	.15	.25
18	S-Conceptual Span	.10	.11	.26	.26	.29	.31	.26	.22	.25	.32	.44	.38	.26	.30	.20	.15
19	S-Listening Sentence	.07	.06	.15	.17	.12	.14	.17	.12	.13	.18	.30	.23	.21	.14	.24	.05
20	S-Updating	.12	.14	.33	.26	.32	.31	.28	.15	.23	.16	.25	.33	.33	.14	.18	.26
21	S-Rhyming	.21	.23	.39	.34	.34	.36	.32	.26	.27	.18	.21	.26	.36	.19	.25	.25
22	E-Forward Digit Span	.21	.25	.40	.38	.24	.24	.35	.32	.35	.25	.12	.13	.32	.23	.14	.35
23	E-Word Span	.14	.15	.31	.30	.21	.18	.32	.28	.34	.17	.12	.06	.25	.21	.13	.28
24	E-Phonetics Span	.13	.13	.25	.23	.18	.23	.27	.21	.26	.23	.15	.11	.27	.22	.08	.21
25	E-Backward Digit Span	.19	.25	.49	.46	.37	.33	.41	.29	.32	.15	.19	.19	.42	.27	.20	.33
26	S-Forward Digit Span	.20	.24	.34	.30	.34	.37	.27	.17	.22	.24	.28	.30	.31	.13	.10	.23

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
27	S-Backward Digit Span	.25	.28	.41	.41	.46	.40	.34	.26	.26	.27	.39	.37	.41	.24	.19	.24
28	S-Word Span	.13	.13	.21	.23	.24	.27	.23	.17	.17	.30	.33	.30	.21	.21	.09	.17
29	S-Phonetics Span	.00	.04	.14	.12	.16	.15	.23	.17	.20	.17	.15	.14	.15	.24	.10	.07
30	Visual Matrix	.06	.10	.34	.38	.26	.21	.35	.28	.32	.21	.13	.19	.35	.22	.18	.20
31	Mapping-Directions	-.03	-.02	.13	.08	.13	.11	.09	.05	.06	.02	.10	.10	.12	.06	.04	-.01

	Variable	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
17	E-Rhyming	--														
18	S-Conceptual Span	.09	--													
19	S-Listening Sentence	.04	.20	--												
20	S-Updating	.29	.15	.06	--											
21	S-Rhyming	.25	.17	.12	.27	--										
22	E-Forward Digit Span	.12	.22	.04	.23	.22	--									
23	E-Word Span	.15	.17	.11	.27	.20	.42	--								
24	E-Phonetics Span	.07	.14	.07	.11	.22	.27	.29	--							
25	E-Backward Digit Span	.21	.20	.09	.25	.26	.40	.25	.17	--						
26	S-Forward Digit Span	.11	.21	.12	.37	.30	.49	.37	.27	.28	--					
27	S-Backward Digit Span	.16	.29	.17	.31	.30	.26	.21	.09	.44	.37	--				
28	S-Word Span	.05	.24	.18	.27	.26	.31	.31	.26	.22	.48	.29	--			
29	S-Phonetics Span	.09	.16	.14	.10	.17	.17	.25	.34	.12	.28	.09	.29	--		
30	Visual Matrix	.11	.17	.09	.20	.17	.18	.14	.14	.25	.08	.30	.11	.11	--	
31	Mapping-Directions	.03	.05	.01	.00	.04	.00	.01	.13	.12	-.04	.07	.02	.09	.09	--

Note. All correlation (*rs* > .10) significant at an alpha of p < .05.

Appendix C

Standardized Estimates Used to Create Factor Scores

Factor Scores	*M*	*SD*	Standardized loading
English reading			
E - word recognition	34.98	11.35	.95
E - comprehension	12.86	5.12	.89
Spanish reading			
S - word recognition	25.15	12.5	.88
S - comprehension	7.02	4.03	.88
English oral language			
E – expressive vocabulary	47.9	13.55	.75
E – receptive vocabulary	97.46	20.03	.84
E - syntax	8.51	5.89	.85
Spanish oral language			
S – expressive vocabulary	28.20	16.71	.79
S – receptive vocabulary	45.06	16.52	.69
S - syntax	7.93	6.69	.55
English STM			
E – digit forward	3.55	0.91	.66
E – digit backward	2.08	0.95	.59
E – word span	2.32	0.78	.56
E – pseudo word span	1.20	0.69	.40
Spanish STM			
S – digit forward	3.35	0.82	.68
S – digit backward	1.76	0.97	.57
S – word span	1.78	0.84	.57
S – pseudo word span	1.18	0.77	.33
English WM (executive)			
E – conceptual span	3.08	3.08	.58
E – listening sentence span	1.42	0.41	.47
E - updating	2.19	1.90	.55
E – rhyming span	3.30	3.26	.41
Spanish WM (executive)			
S – conceptual span	1.63	0.48	.47
S – listening sentence span	1.26	0.30	.27
S - updating	1.47	0.58	.45
S – rhyming span	2.96	2.53	.45
Visual – Spatial Sketchpad			
Matrix	11.14	6.69	.63
Mapping & Direction	1.40	0.55	.16

Note. E = English; S = Spanish; WM = Working Memory; STM = short-term memory or phonological loop.