Does the Structure of Working Memory in EL Children Vary Across Age and Two Language Systems?

H. Lee Swanson

University of New Mexico

Milagros Fatima Kudo

University of California, Riverside

M. Lee Van Horn

University of New Mexico

Authors Note

H. Lee Swanson, Area of Educational Psychology, University of New Mexico (HLswanson@unm.edu) and University of California-Riverside (lee.swanson@ucr.edu), Milagros Kudo (MilyKudo@gmail.com), Area of Educational Psychology, University of California-Riverside, and M. Lee Van Horn (MLVH@unm.edu), Educational Psychology-Methodology Group, University of New Mexico.

The data for the major part of this study is based on a four-year longitudinal study funded by the California Linguistic Minority Research Institute Grant (Grant 02002CY-01CG-R) and the U.S. Department of Education, Cognition and Student Learning (USDE R324A090092) and completion of this work was supported by an NSF grant (DRL 1660828) awarded to the first author. Special appreciation is given to, Jill Leafstedt, Kelly Roston, Danielle Guzman-Orth, Joseph Rios, Elizabeth Arellano, Nicole Garcia, Alfredo Aviles, Steve Gómez, Paula Aisemberg, Valerie Perry, Loren Albeg, Dennis Sisco-Taylor, Wenson Fung, and School District Liaison and Consultant: Erin Bostick Mason, Leilani Sáez and Cathy Lussier for data collection and/or analysis.

Correspondence should be sent to H. Lee Swanson, Ph.D., Area of Educational Psychology, College of Education/COE Research, Albuquerque, New Mexico, 87131. HLswanson@unm.edu.

Abstract

This study examined the cross-sectional structure of working memory (WM) among elementary school English learners (ELs). A battery of WM tasks was administered in Spanish (L1) and English (L2) within five age groups (ages 6, 7, 8, 9, & 10). Confirmatory factor analysis showed a three-factor structure of WM emerged in both L1 and L2 administrations for each age group. The important findings, however, were: (1) the separation between the executive component and storage component (phonological loop) structure of WM increased as a function of age within both language systems, (2) the structure of WM supported a domain general phonological storage component and a domain general executive system across both language systems, and (3) the visual-spatial WM system shared minimal variance with the executive system. Taken together, the findings support Baddeley's multicomponent model (e.g., Baddeley & Logie, 1999) as a good fit to the structure of WM in EL children's English and Spanish language system.

Keywords: working memory, bilingual children, multicomponent model

Does the Structure of Working Memory in EL Children Vary Across Age and Two Language Systems?

Working memory (WM) is an important factor in children's classroom learning. In general, children with higher levels of WM proficiency tend to do well in reading, mathematics, and other subjects (Swanson & Alloway, 2011, for review). Despite knowing the importance of WM to learning, little research has included English language learners (ELs). This is surprising considering the number of ELs in the U.S. is increasing. In fact, between 1980 and 2010 the number of school-age children (5-17 year olds) who spoke a second language at home rose from 4.7 to 11.2 million; in other words, it rose from 10% to 21% for the population in this age range (U.S. Department of Education, National Center for Education Statistics, 2012). However, school achievement tends to be lower for ELs who speak Spanish as their first language than other minority groups (McCardle, Keller-Allen, & Shuy, 2008). Given this information, learning about cognitive processes, such as WM, in children exposed to more than one language is essential, especially given its impact on academics.

One framework to capture diverse memory processes as they apply to EL children is Baddeley's multicomponent WM model (Baddeley, 2012; Baddeley & Logie, 1999). This multicomponent model characterizes WM as comprising a central executive controlling system that interacts with a set of two subsidiary storage systems: the speech-based phonological loop and the visual-spatial sketchpad. The phonological loop is responsible for the temporary storage of verbal information; items are held within a phonological store and are maintained within the store through the process of subvocal articulation. The phonological loop is commonly associated with short-term memory (STM) because it involves two major components discussed in the STM literature: a speech-based phonological input store and a rehearsal process (see Baddeley & Logie, 1999, for a review). The visual-spatial sketchpad is responsible for the storage of visual-spatial information over brief periods and plays a key role in the generation and manipulation of mental images. The central executive is involved in the control and regulation of the WM system. According to Baddeley (Baddeley, 2012; Baddeley & Logie, 1999), the central executive coordinates the two systems, focusing and switching attention, and activating representations within long-term memory (LTM). The central executive is thought to play an important role in "controlled attention," which coincides with Norman and Shallice's (1986) Supervisory Attentional System (SAS) model. This model has been revised to include an episodic buffer (Baddeley, 2000), but support for the tripartite model has been found across various age groups of children (e.g., Gathercole, Pickering, Ambridge, & Wearing, 2004). For example, Gray et al. (2017) added measures of the episodic component (tasks that bind verbal and visual information) to a battery of WM measures to 7–9-year-olds and found weak support for the four factor model when compared to Baddeley's (e.g., Baddeley & Logie, 1999) earlier three factor model. Thus, this study will focus on the three factor structure consistent with Baddeley's earlier model.

Although a plethora of studies have supported the three-factor structure of WM in monolingual children, a question emerges as to whether the same WM structure occurs in EL children. This question is raised for two reasons. The first reason is that the executive system may not operate independent of the phonological system. Because EL children may experience difficulties navigating between two language systems, a differentiation between the two processes may not be possible. This hypothesis appears reasonable because STM storage and WM have not been consistently differentiated, even in monolingual children. For example, Hutton and Towse (2001) found that both WM and STM tasks loaded on the same factor for children 8 to 11 years old. Their results also showed that correlations related to WM and STM on measures of reading were of the same magnitude (see Table 4), suggesting that WM and STM share the same construct (also see Cowan, Towse, Hamilton Saults, Elliot, Lacey, Moreno, & Hitch, 2003, for a similar finding). As stated by Hutton and Towse (2001), "…it appears that what holds for WM in adults may not be equally true for children, and vice versa. The study highlights the value of taking account of children's online processing during WM tasks, and in so doing suggests that WM and STM may, at least in some circumstances, be rather equivalent" (p. 392). These findings, however, may be contrasted with Gathercole et al. (2004) who found support for the basic modular structure of WM in children that was consistent with a tripartite adult-based model of Baddeley (i.e., phonological STM, executive processing, and the visual-spatial sketchpad). However, it was unclear from Gathercole et al.'s results whether the processes that contributed to these structures were distinct. For example, the average factor correlation between measures of STM (phonological loop) and executive processing was .80 (p. 186). In addition, the magnitude of this correlation varied across age groups suggesting that the factor structure was less distinct for the younger participants than the older participants. The authors indicated that the close association between the factors associated with the phonological STM and the executive component of WM was because "the central executive's identification was based on tasks that are constrained by phonological loop capacity" (p. 188). Given these findings with monolingual children, further work is necessary to determine if EL children's STM (phonological loop) is distinct from the central executive component of WM.

Second, the research is unclear as to whether visual-spatial WM operates independent of a domain general or domain specific (verbal) system. Bialystok and Feng (2009) noted that bilinguals consistently outperform monolinguals in all nonverbal measures of attention and control; however, this bilingual advantage does not extend to tasks of language processing (e.g., phonological loop or STM). One explanation for this outcome has been attributed to reduced verbal abilities of bilingual children (Bialystok, 2009). That is, the weak relationship between WM and bilingualism emerges because bilinguals' lexical knowledge is distributed across all their languages, and therefore do not have the full range of language proficiency when compared to their monolingual peers (e.g., Namazi & Thordardottir, 2010). Thus, there is the assumption that visual-spatial WM may not share resources with verbal measures and/or may operate independent of a domain general system. This finding contrasts a popular model that suggests a domain general component manages both the verbal and visual spatial WM information in children (Baddeley & Logie, 1999; Engle et al., 1999). Thus, it is possible there is a functional dissociation between the verbal and visual systems in EL children.

The purpose of this study was to determine the structure of WM in English language learners. One important issue addressed by the study is whether the structure of WM is consistent across age groups of EL children, or is subject to changes as a function of developmental growth. One possibility we explore is that phonological storage and executive processing are distinct components in older children when compared to younger children. There is some literature suggesting that growth of the executive component of WM is unstable across age and that STM (i.e., phonological loop) shows a age-related stable increase (Colom, Rebollo, Abad, & Shih, 2006; Shahabi, Abad, & Colom, 2014). Thus, the strength of the relationship between the executive component WM and storage components of WM may change across age. Specifically, the research questions addressed here include:

1. Does the theoretical framework for working memory (WM) in English and Spanish fit the data for children who are English language learners (ELs)?
2. Is there an age-related increase in visual-spatial processing dominance relative to phonological and executive components of WM?
3. Is there an age-related increase towards independence between the phonological and executive system?

To address the above questions, a confirmatory factor analysis (CFA) was conducted within and across the English and Spanish system as a function of each age group. As mentioned, the Baddeley and Logie (1999) WM model reflects a multi-process activity in which capacities are allocated over a variety of systems. Of course there are competing models. Some models assume that WM allows for the simultaneous processing of a specialized pool of resources that overlap on a single factor, such as controlled or focused attention (e.g., Chow & Conway, 2015; Cowan, Saults & Blume, 2014), whereas other models suggest there are specialized pools of resources that reflect isolated capacity pools and operations not shared with a single factor (e.g., Shah & Miyake, 1996; Mammeralla, Borella, Pastore & Pazzaglia, 2013). Some models also describe long-term memory and WM as separate but related aspects of one system (e.g., Unsworth, 2010). Others, such as Cowan (1995), do not view WM as structurally different from long-term memory, but WM consists of automatically and attentionally activated information. Others view the content of WM as active long-term memory representations (e.g., Ericsson & Kintsch, 1995).

To simply our comparison to the Baddeley and Logie (1999) model, three competing models were tested. Model 1 (as shown in Figure 1) includes one factor that assumed all WM measures load on the same factor. Given the findings that STM and WM task are difficult to distinguish in young children (e.g., Hutton & Towse, 2001), it is possible that a multicomponent model of WM may not occur in either an English or Spanish language system. A two-factor model was tested that compared storage and controlled attention factors (see Figure 2). The assumption was that a two-factor model with distinct storage systems could be separated from WM (Engle et al., 1999). That is, simple span tasks (STM) load on a separate factor from complex span tasks (WM).). Although WM or complex span tasks share the same processes (e.g., rehearsal, updating, controlled search) as short-term memory (STM) or simple span tasks, simple tasks (e.g., recalling words or digits in the order of presentation) have a greater reliance on phonological processes than WM or complex span tasks (e.g., recalling words or digits in the context of interference/distraction) (see Unsworth & Engle, 2007, pp. 1045-1046, for a review). Consistent with the aforementioned Baddeley and Logie (1999), a three-factor model (see Figure 3) was also tested that included the central executive, phonological loop, and the visual-spatial sketch pad.

Methods

Participants

Data for this study were derived from two larger federally-funded studies investigating WM (Authors). Children from Grades 1 through 5 were tested and ages are reported in Table 1. Also reported are the scores for vocabulary and reading performance on standardized norm-referenced measures. A total of 614 students in Grades 1 through 5 from four large school districts in the U. S. southwest were included in this study. All children were Hispanic and exposed to both English and Spanish languages. The first language for all children participating in this study was Spanish. Parent interviews indicated that in 82% of households, the children's primary current home spoken language was Spanish. All children participating in the study were designated as English language learners (ELs) based on school administration of the California English Language Development Test (CELDT).

Table 1 reports the percentages of children on a federal lunch program and percentage of time that Spanish is spoken in the home, and mean standard scores on English and Spanish measures of vocabulary and reading. Also reported are the normed-referenced scores on a measure of fluid intelligence (Raven Colored Progressive Matrices Test; Raven, 1976). As noted, fluid intelligence and reading were in the normal range for this sample, whereas vocabulary scores were in the low average range. For this study, although these children would be considered bilingual, there is heterogeneity within the existing EL classification. Table 1 provides the normed reference scores for each cohort. As shown for the total sample, mean English word identification, fluid intelligence, and math scores were in the average range. Spanish scores, especially passage comprehension, were substantially lower than English and were outside the normal range, suggesting that the sample was not biliterate.

Classification Measures

Vocabulary-English. The Peabody Picture Vocabulary Test (PPVT-III; Dunn & Dunn, 1981) was administered to assess English receptive vocabulary knowledge. Children were presented with four pictures and were asked, after hearing a word spoken in isolation, to select the picture that matched the meaning of the word. The technical manual states a parallel form reliability of .91.

Vocabulary-Spanish. The Test de Vocabulario en Imagenes (TVIP) is similar to the PPVT-III in the presentation and administration (Dunn, Lugo, Padilla, & Dunn, 1986). Children were presented with four pictures and asked to identify the picture for a word read aloud in Spanish. The split-half reliability presented in the manual was .91 to .94.

Reading-English and Spanish. The Woodcock-Muñoz Language Survey-Revised (WMLS-R) Spanish and English word identification and passage comprehension tests were administered to establish a normed-referenced reading level in English and Spanish (Woodcock, 1998; Woodcock, Muñoz-Sandoval, & Alverado, 2005). The test reliabilities range from the mid-.70s to high-.90s for the word identification test and the high-.80s to high-.90s for the passage comprehension test in the various age clusters.

Fluid (nonverbal) intelligence. It was necessary to ensure that the sample did not have low general intellectual performance. Thus, the Raven Colored Progressive Matrices (CPM; Raven, 1976) was used as an indicator of nonverbal or fluid intelligence. Children were given a booklet with patterns displayed on each page, which revealed a missing piece. For each pattern, six possible replacement pieces were displayed. The dependent measure was the number of matrices solved correctly. The technical manual reports internal consistency reliability ratings ranging from .80 to .90.

Working Memory Measures

Short-term memory (STM) measures (phonological loop). Three measures of STM were administered in Spanish and English: Forward Digit Span, Word Span, and Pseudoword Span. The Forward Digit Span task (taken from the WISC-III; Wechsler, 1991) and a Spanish translated version were administered. The Forward Digit Span task required children to recall sequentially ordered sets of digits that increased in set size, which were spoken by the examiner. The technical manual reported a test-retest reliability of .91. The dependent measures was the largest set of items recalled in order (range = 0 to 8). For the translated Spanish version of the Digit Span subtest, identical numbers were presented in the same order as the English version. There were no deviations in procedure, except for language use.

The Word Span and Pseudoword Span tasks were presented in the same manner as the Forward Digit Span task. In the Word Span task examiners read lists of one- or two-syllable, high frequency words that included unrelated nouns and then asked the children to recall the words. Word lists gradually increased in set size, from a minimum of two words to a maximum of eight. The Pseudoword Span task (Phonetic Memory Span task) uses strings of one-syllable nonsense words, which are presented one at a time in sets of 2 to 6 nonwords (e.g., DES, SEEG, SEG, GEEZ, DEEZ, DEZ). A parallel version was developed in Spanish for the Word Span and Pseudoword Span tests. The dependent measure for all STM measures was the highest set of items retrieved in the correct serial order (range = 0 to 7).

Executive component of WM. A Conceptual Span, Listening Sentence Span, and Updating task were administered in English and Spanish to capture the executive component of WM. Previous studies have shown that these measures load on the executive component of WM (see Swanson, 2008). The WM tasks require children to hold increasingly complex information in memory while simultaneously responding to a question about the task. For example, after children listened to a list of words they were asked, “Which word from the list did I say, X or Y?” They were then asked to recall words from the list. This balance of simultaneous storage and processing is consistent with a number of studies of WM processing, including Daneman and Carpenter's (1980) seminal WM measure. A previous study (Swanson, 1996) with a different sample, established the reliability and the construct validity of the WM measures with the Daneman and Carpenter measure.

The Conceptual Span task (Swanson, 2008) was used as an indicator of WM processing that involves the ability to organize sequences of words into abstract categories. Children listened to a set of words that, when re-organized, could be grouped into meaningful categories.

For example, they were told a word set, such as, “shirt, saw, pants, hammer, shoes, nails.” After answering the distracter question, they were asked to recall the words that “go together” (i.e., shirt, pants, and shoes; saw, hammer, and nails). The range of set difficulty was two categories containing two words each to four categories with four words each. A Spanish-translated version was also administered. Care was taken in the development of the measure to keep the abstract categories the same in both languages (e.g., clothes and tools); however, WM-level appropriate words were used in cases where direct translation resulted in significantly harder words to recall. The dependent measure for both versions was the number of sets recalled correctly (range = 0 to 6).

The children's adaptation of Daneman and Carpenter's (1980) Listening Sentence Span task was administered. This task required the presentation of groups of sentences, read aloud, for which children tried to simultaneously understand the sentence contents and to remember the last word of each sentence. The number of sentences in the group gradually increased from two to six. After each group of sentences was presented, the child answered a question about a sentence and then was asked to recall the last word of each sentence. The dependent measure was the total number of correctly recalled word items in order up to the largest set of items (e.g., set 1 contained 2 items, set 2 contained 3 items, set 3 contained 4 items, etc.), in which the process question was also answered correctly.

Because WM tasks were assumed to tap a measure of controlled attention referred to as updating (e.g., Miyake, Friedman, Emerson, Witzki, & Howerter, 2000), an experimental Updating task, adapted from Swanson et al. (2004), was also administered. A series of one-digit numbers was presented that varied in set length. No digit appeared twice in the same set. The examiner told the child that the length of each list of numbers might be 3, 5, 7, or 9 digits.

Children were then told that they should only recall the last three numbers presented. Each digit was presented at approximately one-second intervals. After the last digit was presented the child 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. Two measures were administered to assess visual-spatial WM: Visual Matrix and Mapping & Directions tasks (Swanson, 2008). The Visual Matrix task assessed the ability of participants to remember visual sequences within a matrix. Participants were presented a series of dots in a matrix and were allowed 5 seconds to study the matrix. The matrix was then removed and participants were asked, in both English and Spanish, "Are there any dots in the first column?" To ensure the understanding of columns prior to test, participants were shown the first column location and then practiced finding it on blank matrices. In addition, for each test item the experimenter pointed to the first column on a blank matrix (a grid with no dots) as a reminder of first column location. After answering the discrimination question, students were asked to draw the dots they remembered seeing in the corresponding boxes of their blank matrix response booklet. The task difficulty ranged from a matrix of 4 squares and 2 dots to a matrix of 45 squares and 12 dots. The dependent measure was the number of matrices recalled correctly (range = 0 to 11).

The Mapping and Directions task required children to remember a sequence of directions on a map. The experimenter presented a street map with dots connected by lines; arrows illustrated the direction a bicycle would go to follow this route through the city. The dots represented stoplights, while lines and arrows mapped the route through the city. The child was allowed 10 seconds to study the map. After the map was removed, the child was asked a process question (i.e., "Were there any stop lights on the first street (column)?"). The child was then presented a blank matrix on which to draw the street directions (lines and arrows) and stop lights (dots). Difficulty ranged on this subtest from 4 dots to 19 dots. The dependent measure was the highest set of correctly drawn maps (range = 0 to 9), in which the distracter process question was also answered correctly.

Procedures

Children were tested individually after informed consent was obtained for participation. For each cohort, two sessions of individual testing were conducted, each lasting thirty minutes (for limited English or Spanish speakers) to one hour. The presentation order of tasks was counterbalanced into one of six presentation orders. Children were randomly assigned to each participation order and randomly assigned to an examiner. No Spanish and English versions of the same test were presented consecutively.

Statistical Analyses

To analyze whether the data from EL students fit the proposed Baddeley model of WM, a series of confirmatory factor analyses (CFAs) were conducted. The proposed model structure included 3 latent factors, composed of the central executive components of WM (which includes conceptual span, listening span, and updating tasks), the phonological loop (which includes forward digit span, word span, and pseudoword span tasks), and the visual-spatial sketchpad (which includes mapping & directions span and visual matrix tasks; Figure 3; Swanson, 1993). All the analyses were conducted in SAS, PROC CALIS, version 9.4 (SAS, 2012).

Several indices were used to test the structure of the WM across the age groups. The Comparative Fit Index (CFI; Bentler, 1990) was the primary index for our analysis. Values for CFI range from 0 to 1.00 and are derived from comparisons between the hypothesis and independence models. Values > .97 are considered a good fit to the data (Schermelleh-Engel, Moosbrugger & Müller, 2003), whereas Hu and Bentler (1999) suggest a CFI > .94 as an acceptable fit. In general, CFI's in the mid .90s or above are associated with models that are plausible approximations of the data. Also reported is the root-mean-square error of approximation (RMSEA) and the standardized root square residual (SRMR). An RMSEA below .05 with the left endpoint of its 90% confidence interval that is markedly smaller than .05 represents an excellent fit (e.g., Raykov & Marcoulides, 2006). The SRMR represents the average residual value derived from fitting the variance-covariance matrix for the hypothesized model to the variance-covariance matrix of the sample data. Values range from 0 to 1.00; in a well-fitting model, an SRMR is .05 or less. To compare competing or alternative models (one factor vs. two vs. three factors), the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) were used. Models with smaller values are preferred to models with higher values when determining the best model fit. The AIC is primarily focused on comparing competing nonhierarchical models (Kline, 2011), and the BIC is recommended when the sample size is large and the number of parameters is small. Further, the BIC is more likely to penalize for additional model parameters than the AIC.

Results

The means, standard deviations, correlation matrices, and gender representations for each age group are reported in Appendix A and B. Within each age group, we determined which measures met standard criteria for univariate analysis (Kline, 2011). Mardia's normalized coefficient of multivariate normality was checked to ensure that the data met the assumptions of Maximum Likelihood estimation. No clear outliers (e.g., > 3.5 SDs from the mean) were identified. Skewness for measures less than 3 and kurtosis less than 4 occurred for some measures within age groups. Transformations were conducted on these variables and the results of each model with and without these transformations were compared. Since there were no substantial differences in the model results, the untransformed scores (except for the listening span measure) were used for analyses. The listening span measure was difficult for the younger aging and therefore skewing the data. We will now address three questions raised in the introduction.

Question 1: Does the theoretical framework for working memory (WM) in English and Spanish fit the data for children who are English language learners (ELs)?

The fit indices for the one-, two- and three-factor models within each age group are shown in Table 2. Although the model fit of the data generally supported a three-component model for most age groups, some age groups provided a better fit to the model than others. As shown in Table 2, the three-factor model provided a good fit for four of the five groups within the English and Spanish language system. A poor fit was found for the 7-year-old age group on the English measures, and a poor fit was found for the 9-year-old group on the Spanish measures.

To investigate the above variations in model fit, we analyzed the role of gender and span across the measures. Previous studies have found that variations in WM occur as a function of gender (e.g., León, Cimadevilla, & Tascón, 2014; Piccardi, Leonzi, D'Amico, Marano, & Guariglia, 2014), and that WM scores are higher in the first rather than the second language system (e.g., Thorn & Gathercole, 1999). Thus, we computed a 2 (gender) x 5 (English and Spanish measures for STM, WM, and the visual-spatial sketch pad) MANCOVA on the span scores within each age group. The covariate was the composite score for English and Spanish vocabulary. The mean vocabulary scores and the z-score for each of the five latent measures are shown in Table 3 as a function of gender and span measures. The z-scores for each factor were based on the mean and SD for the total sample.

The *F* ratios from the MANCOVA for each comparison are shown at the bottom of Table 3. As shown, a significant gender effect occurred for age 7 in which an advantage was found for females relative to male performance, especially on the Spanish STM latent measure. An interaction occurred for the 7 year old age group which indicated low performance for males on the Spanish STM measure relative to other memory measures.

A significant effect related to span scores occurred across all age groups. A Tukey test yielded significant differences (*p*s < .05) in span scores for visual-spatial (Vis), English & Spanish STM (E-STM & S-STM, respectively), and English & Spanish WM (E-WM & S-WM, respectively) for age 6 (S-STM = S-WM = Vis > E-STM = E-WM), age 7 (E-STM = E-WM = Vis > S-STM = S-WM), age 8 (S-STM = E-STM = E-WM = Vis > S-WM), age 9 (Vis > S-WM = E-WM = E-STM > S-STM), and age 10 (Vis > E-STM = S-WM = E-WM > S-STM). The general pattern for span scores was an increasing advantage for visual-spatial measures when compared to the verbal STM and verbal WM measures with increases in age.

Question 2: Is there an age-related increase in visual-spatial processing dominance relative to phonological and executive components of WM?

As shown in Table 3, and as expected, there is an age-related increase in performance across all span scores. Particularly noteworthy was the increase in visual-spatial WM. This is a common finding related to a bilingual samples where a clear advantage is found for visual-spatial measures relative verbal WM measures. A question emerges as to whether the strength of the correlations between verbal and visual measures becomes increasingly distinct across the age ranges. As shown in Table 4, the correlations among verbal measures with visual-spatial measures were weak across all age groups. Using Cohen's standard for effect sizes (ESs; 1988), the results show that the magnitude of the correlations with the visual spatial span latent measures was small (majority less than .10) relative to the verbal measures. In contrast, correlations between STM and WM across the language system were moderate (.30 range) to large in magnitude (.50 range). Using a large effect size as a criterion (*d* = .50), the results show increasingly stronger cross-language effects for the executive components for the older age groups (ages 9 and 10) than younger age groups (ages 6, 7, and 8). Thus a question emerges as to whether a general language system is in operation across both language systems.

Question 3: Is there an age-related increasing independence between the phonological and executive system?

This question will be first answered within each age group and then between each age group. A bi-factor model was computed that included both English and Spanish measures. The parameter estimates for the first-order factors (specific factors) and the second-order factor (g-factor) for the bi-factor model are shown in Table 5. The three specific factors were: English and Spanish phonological loop measures, English and Spanish executive processing measures, and visual-spatial WM measures. Of interest was whether there was an age-related increase in the loadings on the executive system factor. As shown, the g-factor loadings for measures that tapped the phonological loop yield moderate to high ESs across age on the majority of measures. All estimates yielded moderate to high ESs on the STM measures for the g-factor across all age groups. In contrast, the magnitude of parameters for the executive components loading on the g-factor increased incrementally across age group. That is, the frequency of moderate and high ES (.30 to > .50) for the executive system on the g factor was 1, 2, 4, 4, and 6 for ages 6 to 10, respectively. The Table also shows more frequent moderate to high ESs (.30 to .50) occurred for parameter estimates for the g-factor for the verbal WM measures when compared to the WM visual-spatial measures.

In general, the results provide partial support is found for the notion that the executive system becomes increasingly domain general as a function of age. Although the above analysis implicates a general system across Spanish and English measures, partial support is emphasized because it is unclear as to whether the three factor structure holds within each language system across age groups. The fit indices for three-factor models are shown in Table 6. Table 6 shows that the Baddeley model provided a good fit across all age groups. We determine if this model could be improved upon. We modified the model by dividing the STM measures into English and Spanish storage and combining the executive processes so that they reflected a common system. This modification was based on Kaushanskaya and Yoo (2013) findings which suggested that cross-language associations were stronger for WM tasks than for STM tasks. As shown in Table 6, except for age 8, the BIC indices (lower values indicate a better fit) were lower for the four-factor model fit when compared to the three-factor model. It is important to note, however, a good fit (i.e., CFI, RMSEA & SRMR) was found for the four factor model across all age groups.

However, a statistical comparison was made between the three and four factor model to determine the most parsimonious account of the data. The AIC values (shown in Table 6) were compared to determine if the three factor model provided a more parsimonious model than the four factor model (see Kline, 2011, for rationale). The four factor Model indicated a significantly better fit to the data than the three factor model for age 6, χ^2 (1) = 5.49 (170.38 –164.89), $p < .05$. However, none of the other age groups yielded a significantly better fit for the four factor model when compared to the three factor model (all ps > .05). These findings suggest the three factor models holds across the majority of age groups.

Measurement Invariance

The important question not addressed in the above analyses was whether the model parameters have the same meaning across the age groups within and across each language system. Thus, we tested if the same CFA model between each age group occurred within each language system. The least restrictive comparison (referred to as Configural invariance) was computed to determine if the number of factors and correspondence between the factors and indicators were the same within each language. We then determined if the between groups factor loadings were equivalent in a more restrictive invariance model (referred to as pattern invariance, weak invariance, or metric invariance). This hypothesis is referred to as pattern invariance in this study since we are concerned with equivalence in factor loadings. The restrictive hypothesis determines if the factor loadings are manifested in the same way for each age group. Cheung and Rensvolt (2002) indicated that changes in CFI values (Δ CFI) less than or equal to .01 indicate that the stricter invariance hypothesis should not be rejected. Further testing for the invariance hypothesis included a Chi-square difference test. A non-significant difference in the Chi-square between the convergence and pattern invariance hypothesis suggested that the stricter non invariance test should not be rejected.

As shown in Table 7, the stricter invariance hypothesis (factor loadings between groups were equivalent) was supported (Δ CFI) for all English measure except when comparing group 7 and 9 year olds (Δ CFI > .01). In contrast, the majority of comparisons of memory measures administered in Spanish did not support the stricter invariance hypothesis suggesting that substantial variations in factor loadings emerged between age groups on the Spanish measures. This analysis was interpreted as suggesting there was an age related distinction related to the WM measures as a function of language. Based on the results shown in Table 5 (STM measures consistently load on a g-factor across age whereas the WM measures shown an age-related increment in g loading), we tested whether the factor loadings varied across age groups. That is, our assumption was that independence between phonological and executive systems was more likely to occur at the older than younger ages, suggesting the invariance hypothesis related between factor loadings would no hold when comparing younger with older ages. To directly investigate this hypothesis, a multiple-sample comparison was made between age groups when combining performance on the English and Spanish memory measures. Consistent with the notion that executive processing is domain general, one factor included the administration of English and Spanish for measures of listening span and conceptual span. Updating was eliminated since it loaded on both STM and WM factors. A second and third factor was created to capture the phonological loop. According to Baddeley et al. (p170), “the phonological loop has been attributed to “recall of unrelated words, especially digit span. We propose that the primary function of the phonological loop is the processing of novel speech input”. Thus the model we next tested is in line with Baddeley et al. (1998) assumption that that phonological loop reflects a storage system that is not related to familiar items. A distinction was made between phonological loop in terms of items that have access to LTM (digits, words) and a system in which has a weak connection to lexical information in LTM (nonwords). The model is tested by comparing a three factor model that included an language general executive system factor (English and Spanish measures of listening span, conceptual span), a phonological loop factor that reflected familiar or LTM representations of English and Spanish items (digit span, word span) and a phonological factor that reflect storage of unfamiliar items (English and Spanish pseudo words). This model was first tested within the age groups. The CFI was .91, .91, 1.0, 1.0, .98 for ages 6,7,8,9,10, respectively. This analysis was followed by a comparison of the groups in terms of configural invariance and pattern invariance model.

As shown in Table 8, CFI's were greater than .90 for all comparisons on the configural model except between ages 7 and 9. However, when the analysis considered the factor structure in terms of equivalence in factor loadings, four of ten comparisons at the older ages did not support the pattern invariance hypotheses. Although we could have tested further for partial invariance, this did not make sense given that perfect invariance did not occur in the weak model for all comparisons.

In summary, the results suggest that age-related variation occurs in the factor loadings for the executive and phonological loop across age groups. Because the factor model shows a better fit for the older than younger ages, support was found for the notion that a distinction between the executive and phonological loop is greater in the older than younger ages.

Discussion

In summary, the results from the CFAs provide evidence that Baddeley's three-component WM model in English and Spanish is tenable in EL populations for the age groups represented in this sample. More importantly, the cross-sectional results yield three important findings related to development of WM in EL children in the elementary grades. First, there was a trend towards an age-related increase in the cross-language performance on Spanish and English executive processing measures. Although association of the phonological loop with a common system (g factor) was apparent across all age groups, the cross-language association of executive processing measures with a common system did not emerge until age 10. Second, the results found support for the notion that visual WM in this sample operates independent of a domain general system in EL children. The correlational and CFA analyses show independence of visual-WM from the verbal system across all age groups. Finally, the factor structure across both language systems supported a three-factor model that included a phonological loop, executive processing and visual-spatial components of WM. The results related to the questions that directed this study are now reviewed.

Question 1: Does the theoretical framework for working memory (WM) in English and Spanish fit the data for children who are English language learners (ELs)?

To examine whether the theoretical framework of WM, proposed by Baddeley, fits the data for EL students, a series of CFAs were conducted. The results of this study showed that the three-factor structure provided an adequate fit within the two language systems among EL children and, therefore, was comparable to the factor structure for STM and WM found among monolingual children (e.g., Gathercole et al., 2004). The three-factor model had adequate to good fits within each language system for the majority of age groups. When other competing models were tested (a two-factor and a general WM factor model), the model fits were significantly worse, suggesting that the proposed Baddeley model provides the best fit overall.

When generalizing to adult samples (i.e., Engle et al., 1999), we also found that letter, digit, and word span tasks define the STM factor, whereas tasks that include the combination of both processing and storage define the executive component of the WM factor. We defined the processing component of our WM measures as that part of the memory task that directed children to respond to discrimination questions about recently encoded information, whereas we defined the storage component as that part of the task that required children to retrieve recently encoded information. In sum, our findings do not support studies showing processes related WM and STM cannot be differentiated in children (e.g., Hutton & Towse, 2001).

The above findings are important because it is assumed that bilingual children (in this case, EL children) experience some advantage related to the executive system. A number of studies that have shown proficiency in L1 and L2 (i.e., bilingualism) positively influences executive functioning, flexibility, and attentional control (e.g., Bialystok, 2011; however see De Bruin, Treccani, & Della Sala, 2015, for alternative explanation for these findings). Because executive processes are related to WM (Engle, 2002; Friedman et al., 2007), one would expect variations in bilingual proficiency to play an important role in WM performance. However, the majority of studies implicate phonological STM in variations of language acquisition (e.g., Baddeley, Gathercole, & Papagno, 1998; Gathercole, Willis, Emslie, & Baddeley, 1992; Linck, Osthjus, Koeth, & Bunting, 2013), and not the executive system. For example, children's L2 vocabulary is predicted by phonological memory in their L1 (e.g., Engel de Abreau & Gathercole, 2012; Lipka & Siegel, 2007; Swanson, Sáez, Gerber, & Leafsted, 2004; Thorn & Gathercole, 1999). Because STM appears to be more dominant in the early phases of memory performance, these may make the two processes (STM and WM) less distinct. We did not find direct support for this hypothesis.

Question 2: Is there an age-related increase in visual-spatial processing dominance relative to phonological and executive components of WM?

The results demonstrate an increasing age-related dominance in visual-spatial WM span when compared to verbal measures. Likewise, this dominance or independence of the visual WM from verbal WM is apparent in the correlation matrices across all age groups. Thus, consistent with other bilingual studies, we find performance on the visual-spatial WM measures was clearly differentiated from performance on verbal measures. It is important to note that a span advantage for the visual-spatial WM when compared to verbal WM did not occur until the older age groups. Thus, although our analyses across all age groups suggested an independence between verbal and visual-spatial measures, a visual-spatial span advantage relative to verbal measures was not apparent until age 9 and 10.

The above findings for bilingual children vary with popular models in monolingual samples, suggesting that a domain general component manages both the verbal and visual spatial information in children (Baddeley & Logie, 1999; Engle et al., 1999; Roome, Towse, & Jarrold, 2014). Some studies have suggested that a one-factor model captures the shared variance of verbal and visual-spatial information in which there is a common trajectory, at least up until the adolescent years; whereas after the adolescent years (e.g., Alloway, Gathercole, Pickering, 2006), two separate pools of resources emerge (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). Gathercole et al. (2004) reported a tripartiate structure in children between 6 and 15 years of age. However, not all findings indicate clear separability between different WM components. For example, in a very young age group, six and seven-year-olds, performance related to measures of the visual-spatial sketchpad could not be clearly separated from measures related to the central executive (e.g., Gathercole & Pickering, 2000). Regardless, the structure of visual WM functioning was clearly separated from the verbal measures in this EL sample.

Question 3: Is there an age-related increase towards an independence between the phonological and executive system?

Our results show that components of WM (e.g., STM or the phonological loop and the executive system of WM) in EL children are distinguishable within the child's L1 and L2. Although various studies have shown that bilingual children outperform their monolingual counterparts on tasks (e.g., conflict, switching, and flexibility tasks) used to measure executive functioning (Bialystok & Feng, 2009; Bialystok & Shapero, 2005; Carlson & Meltzoff, 2008; Martin-Rhee & Bialystok, 2008), these benefits have not been directly tied to the executive component of WM in ELs. The results show that WM as an executive system and STM as a representation of the phonological loop load on separate factors within each language system.

However, the association of the executive system with a language general system becomes stronger with increasing age. Previous studies (e.g., Kaushanskaya & Yoo, 2013) have shown that bilinguals performed better on the STM tasks than on WM tasks, that L1 STM performance was superior to L2 STM performance, and that correlation analyses between bilinguals' L1 and L2 performance revealed stronger cross-linguistic associations for WM tasks than for STM tasks. Our findings on the factor structure generally coincide with these findings related to the cross-linguistic associations of WM, but not STM. Our findings suggest that WM tasks may engage domain-general central executive processes in bilinguals. However, we did not finding support (except in the youngest age group) that STM skills (phonological loop) depend on language-specific knowledge in the LI and L2.

What are the theoretical implications of our findings? One developmental implication was that the phonological loop has a strong association with a general system even in the younger age groups, whereas the association of the executive system of WM with a general system increases with age. Undoubtedly, there are at least three arguments to be considered related to our interpretation of findings that the emergence of the executive system as a general processing system is age related.

The first argument questions whether the results better reflect changes in a general storage system rather than age-related development in an executive system (e.g., attentional processes related to the coordination of both storage and processing activities). Recall that the format of the WM tasks required the difficulty of the processing questions to remain constant within task conditions, thereby placing a greater emphasis on storage demands. Thus, because we did not separate out storage and processing separately in the analysis, the results may merely reflect “the capacity to temporarily preserve a reliable memory representation of any given information" (Colom et al., 2006, p. 161). However, it is also important to note the process questions for each task served as distracters to item recall because they reflected the recognition of targeted items and items closely related to nontargeted items. Although there are varieties of attention control not captured with our tasks, performance on the executive processing tasks consistently required the monitoring of process and storage demands. Although we cannot directly determine if processing efficiency rather than storage is the locus of age-related performance or if resource switching played a role, we assume, however, there was some executive requirement associated with the combining process and storage operations.

The second argument questions whether a general language system may more appropriately capture the domain general effect than an executive system. If the vocabulary language system is more developed in the older sample than the younger samples, the domain general structure would better capture the older sample than the younger samples. That is, if a language system better captures the domain general effect in older rather than younger participants, then the covariance among the measures should have yielded a different factor structure across the age groups. The results clearly show that the pattern of the factor structure of WM was comparable between older and younger participants.

A third argument is that the processes that clearly mediate age-related differences between the verbal and visual-spatial WM measures have not been identified. For example, it could be that young children, in relation to older children, are less resistant to interference. Another is that speed of processing underlies developmental aspects of WM (Demetriou et al., 2014). Unfortunately, no measures of processing speed were administered in this study. Although speed of processing may reflect a sizable amount of structural capacity in WM in the present study, some work that includes children with reading and math problems suggests that

WM significantly predicts achievement, even after the influence of articulation speed is partialed from the analysis (Swanson & Beebe-Frankenberger, 2004). In accord with these findings, speed of processing (i.e., articulation speed) may be only one component of a general system that underlies age-related performance. However, additional components that include activating new information and sustaining old information should be considered.

Study Limitations

At least four limitations to the current study should be noted. First, although visual-spatial tasks required processing and storage activities, no comparable storage tasks, as with the verbal STM measures (Digit Forward Span, Word Problem Span, and Phonetic Memory Span), were administered. Thus, we are unable to separate controlled attention and visual storage from these measures. Therefore, the visual-spatial WM tasks should be viewed as a combination of both visual-spatial storage and visual-spatial executive processing (cf. Shah & Miyake, 1996). In addition, it is important to note that the latent variable representing the visual-spatial sketchpad only had two manifest variables as indicators. Some statisticians argue that there should be at least three manifest variables for each latent variable in order to correctly estimate the phi matrix (Bollen, 1989).

A second limitation pertains to the unique population that comprises the study's sample. All subjects were recruited in a southwestern state, in generally low SES areas. The sample does not represent the wide diversity of ELs in the U.S. In fact, our sample consists of a fairly homogeneous group of Hispanic students, mainly of Mexican descent, that may reflect the diversity found in Southwest border states in the U.S. Thus, it may be important to consider carefully the generalizability of this study's results to other groups of EL students.

Third, some limitations are related to sample size. While we conducted tests for univariate and multivariate normality as well as outliers, some of the age groups were of relatively small sample sizes. Finally, because of the cross-sectional nature of the design, the study cannot describe within individual changes over time. Although exploring differences between age groups are beneficial in addressing the hypothesis this study, such differences may reflect a number of unmeasured variables.

Conclusion

Because WM has important implications in academics, this study investigated the construct of Baddeley's model of WM in samples of EL students. This analysis of data yielded three important findings: (1) the theoretical framework for WM in English and Spanish fit the data for children who are ELs, (2) the distinction between the phonological loop and executive system in young children becomes stronger with increases in age, and (3) the cross-language structure of measures is stronger for the executive component of WM than for the phonological loop. Taken together, this research finds that Baddeley's model of WM is an adequate conceptualization for ELs and that the construct of WM can be measured within both English and Spanish language systems.

References

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*(6), 1698-1716. doi:10.1111/j.1467-8624.2006.00968.x

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology, 87*(2)*,* 85-106. doi:10.1016/j.jecp.2003.10.002

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423. doi:10.1016/S1364-661(00)01538-2

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29. doi:10.1146/annurev-psych-120710-100422

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language-learning device. *Psychological Review, 105*(1), 158-173. doi:10.1037//0033-295X.105.1.158

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). Cambridge, U.K.: Cambridge University Press. doi:10.1017/CBO9781139174909.005

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin,107*(2), 238-246. doi:10.1037/0033-2909.107.2.238

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(1), 3-11. doi:10.1017/S1366728908003477

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology*, *65*(4), 229-235. doi:10.1037/a0025406

Bialystok, E., & Feng, X. (2009). Language proficiency and executive control in proactive interference: Evidence from monolingual and bilingual children and adults. *Brain and Language, 109*(2-3), 93–100. doi:10.1016/j.bandl.2008.09.001

Bialystok, E., & Shapero, D. (2005). Ambiguous benefits: The effect of bilingualism on reversing ambiguous figures. *Developmental Science, 8*(6), 595-604. doi:10.1111/j.1467-7687.2005.00451.x

Bollen, K. A. (1989). *Structural equation models with latent variables.* New York: John Wiley & Sons.

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

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. doi: 10.1207/S15328007SEM0902_5

Chow, M., & Conway, A. R. A. (2015). The scope and control of attention: Sources of variance in working memory capacity. *Memory & Cognition, 43*(3), 325-339. doi:10.3758/s13421-014-0496-9

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: L. Erlbaum Associates.

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory & Cognition, 34*(1), 158-171. doi:10.3758/BF03193395

Conway, A. R. A., & Engle, R. W. (1996). Individual differences in working memory capacity:

More evidence for a general capacity theory. *Memory, 4*(6), 577-590.

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition, 34*(8), 1754-1768. doi:10.3758/BF03195936

Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and peripheral components of working memory storage. *Journal of Experimental Psychology: General, 143*(5), 1806-1836. doi:10.1037/a0036814

Cowan, N., Saults, J. S., & Clark, K. M. (2015). Exploring age differences in visual Working memory capacity: Is there a contribution of memory for configuration? *Journal of Experimental Child Psychology, 135*, 72-85. doi:10.1016/j.jecp.2015.03.002

Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., Moreno, M. V., & Hitch, G. J. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General*, *132*(1), 113-132. doi:10.1037/0096-3445.132.1.113

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466. doi:10.1016.S0022-5371(80)90312-6

de Bruin, A., Barbara, T., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science, 26(1),* 99-107. doi: 10.1177/0956797614557866

Demetriou, A., Spanoudis, G., Shayer, M., Van der Ven, S., Brydges, C. R., Kroesbergen, E., Podjarny, G., & Swanson, H. L. (2014). Relations between speed, working memory, and intelligence from preschool to adulthood: Structural equation modeling of 14 studies. *Intelligence, 46*, 107-121. doi:10.1016/j.intell.2014.05.013

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.

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

Engle, R. W., & Kane, M. J. (2004). *Executive attention, working memory capacity, and a two-factor theory of cognitive control*. New York, NY: Elsevier Science.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309-331. doi:10.1037/0096-3445.128.3.309

Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211-245.

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. doi:10.1037/0022-0663.92.2.377

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, 34*(4), 177-190. doi:10.1037/0012.1649.40.2.177

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*, *18*(1), 1-16. doi:10.1002/acp.934

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*(5), 887–898. doi:10.1016/S0093

Gray, S., Green, S., Alt, M., Hogan, T., Kuo, T., Brinkley, S., & Cowan, N. (2017). The structure of working memory in young children and its relation to intelligence. *Journal of Memory and Language, 92,* 183-201. doi:10.1016/j.jml.2016.06.004

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. doi:10.1080/10705519909540118

Hutton, U. M. Z., & Towse, J. N. (2001). Short-term memory and working memory as indices of children's cognitive skills. *Memory, 9*(4-6), 383-394. doi:10.1080/09658210042000058

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189-217. doi:10.1037/0096-3445.133.2.189

Kaushanskaya, M., & Yoo, J. (2013). Phonological short-term and working memory in bilinguals' native and second language. *Applied Psycholinguistics, 34*(5), 1005-1037. doi:10.1017/S0142716412000100

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: The Guilford Press.

León, I., Cimadevilla, J. M., & Tascón, L. (2014). Developmental gender differences in children in a virtual spatial memory task. *Neuropsychology, 28*(4), 485-495. doi:10.1037/neu0000054

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, 21*(4), 861-883. doi:10.3758/s13423-013-0565-2

Lipka, O., & Siegel, L. S. (2007). The development of reading skills in children with English as a second language. *Scientific Studies of Reading, 11*(2), 105-131. doi:10.1080/10888430709336555

Mammarella, I. C., Borella, E., Pastore, M., & Pazzaglia, F. (2013). The structure of visuo -spatial memory in adulthood. *Learning and Individual Differences, 25*, 99-110. doi:10.1016/j.lindif.2013.01.014

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11*(1), 81-93. doi:10.1017/S1366728907003227

McCardle, P., Keller-Allen, C., & Shuy, T. (2008). Learning disability identification: How does it apply to English language learners? In E. L. Grigorenko (Ed.), *Educating individuals with disabilities: IDEIA 2004 and beyond* (pp. 137-164). New York, NY: Springer Publishing.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100. doi:10.1006/cogp.1999.0734

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621-640. doi:10.1037//0096-3445.130.4.621

Namazi, M., & Thordardottir, E. (2010). A working memory, not bilingual advantage, in controlled attention. *International Journal of Bilingual Education and Bilingualism, 13*(5), 597-616. doi:10.1080/13670050.2010.488288

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1-18). New York, NY: Springer Publishing.

Piccardi, L., Leonzi, M., D'Amico, S., Marano, A., & Guariglia, C. (2014). Development of navigational working memory: Evidence from 6- to 10-year-old children. *British Journal of Developmental Psychology, 32*(2), 205-217. doi:10.1111/bjdp.12036

Preacher, K.J.. & Merkel, E.C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological Methods, 17,* 1-14. doi: 10.1037/a0026804

Raven, J. C. (1976). *Colored Progressive Matrices*. London, England: H. K. Lewis & Co. Ltd.

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Reuter-Lorenz, P. A., & Jonides, J., (2007). The executive is central to working memory: Insights from age, performance, and task variations. In A. R. A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 250-271). New York, NY: Oxford University Press.

Roome, H. E., Towse, J. N., & Jarrold, C. (2014). How do selective attentional processes contribute to maintenance and recall in children's working memory capacity? *Frontiers in Human Neuroscience, 8*, 1011. doi:10.3389/fnhum.2014.01011

SAS Institute. (2012). *SAS/STAT software: Changes and enhancements through release 9.3*. Cary, NC: Author.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8 (2)*, 23-74.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*(1), 4-27. doi:10.1037/0096-3445.125.1.4

Shahabi, S. R., Abad, F. J., & Colom, R. (2014). Short-term storage is a stable predictor of fluid intelligence whereas working memory capacity and executive function are not: A comprehensive study with Iranian schoolchildren. *Intelligence, 44,* 134-141. doi:10.1016/j.intell.2014.04.004

Swanson, H. L. (1993). Working memory in learning disability subgroups. *Journal of Experimental Child Psychology, 56*(1), 87-114. doi:10.1006/jecp.1993.1027

Swanson, H. L. (1996). Individual and age-related differences in children's working memory. *Memory & Cognition, 24*(1), 70-82. doi:10.3758/BF03197273

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., & Alloway, T. P. (2012). Working memory, learning, and academic achievement. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational* *psychology handbook, Volume 1: Theories, constructs, and critical issues* (pp. 327-366). Washington, D.C.: American Psychological Association.

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*(3), 471-491. doi:10.1037/002-0663.96.3.471

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

Thorn, A. S., & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology: Section A*, *52*(2), 303-324. doi:10.1080/713755823

Unsworth, N. (2015). Consistency of attentional control as an important cognitive trait: A latent variable analysis. *Intelligence, 49*, 110-128. doi:10.1016/j.intell.2015.01.005

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: The role of controlled search. *Memory & Cognition, 41*(2), 242-254. doi:10.3758/s13421-012-0261-x

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*(6), 1038-1066. doi:10.1037/0033-2909.133.6.1038

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*(4), 392-406. doi:10.1016/j.jml.2010.02.001

U.S. Department of Education, National Center for Education Statistics. (2012). *Trends in academic progress.* Washington, D.C.: Author. Retrieved from NECS website: http://nces.ed.gov/nationsreportcard/subject/publications/main2012/pdf/2013456.pdf

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

Woodcock, R. W. (1998). *Woodcock Reading Mastery Test- Revised* (Form G). Circle Pines, MN: American Guidance Service, Inc.

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

Figure 1. One Factor Model of Working Memory (Model 1).

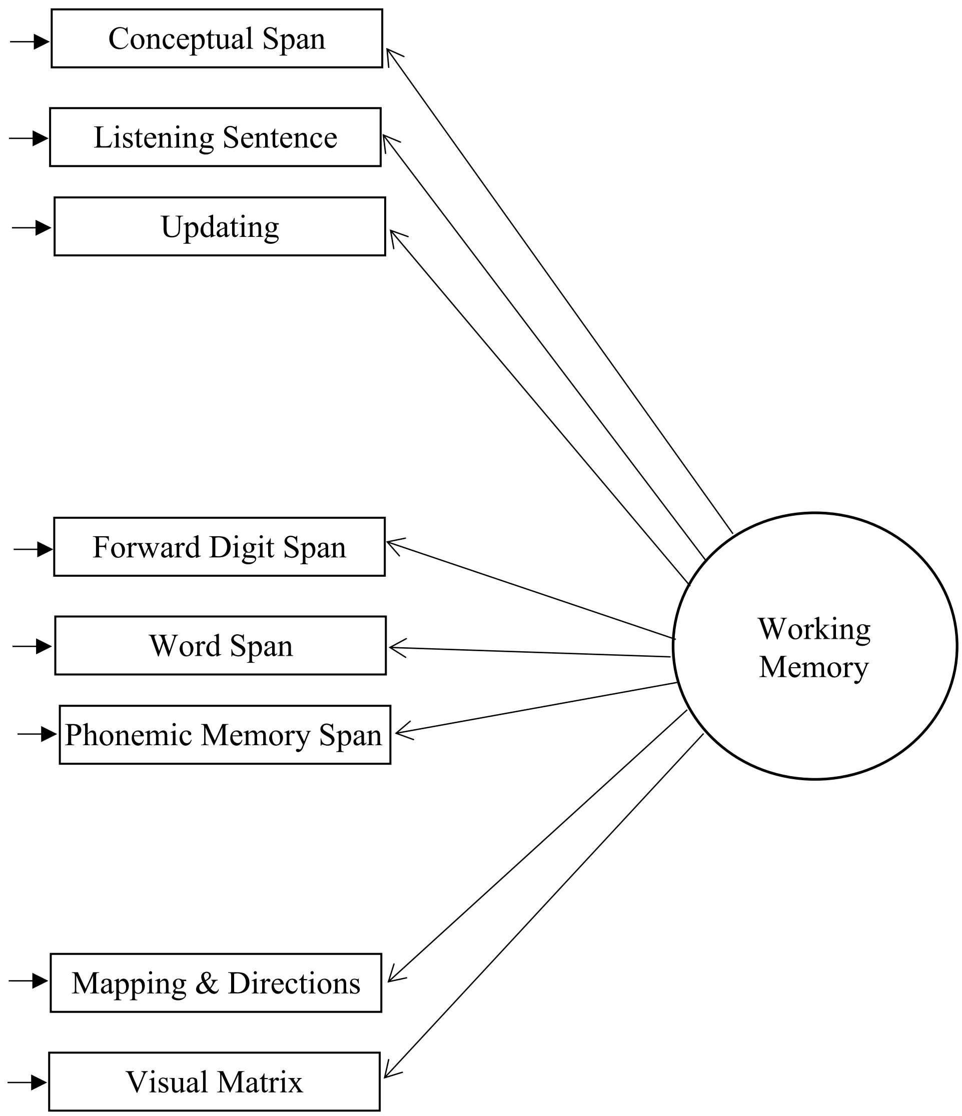

Figure 2. Two-Factor Model of Working Memory (Model 2).

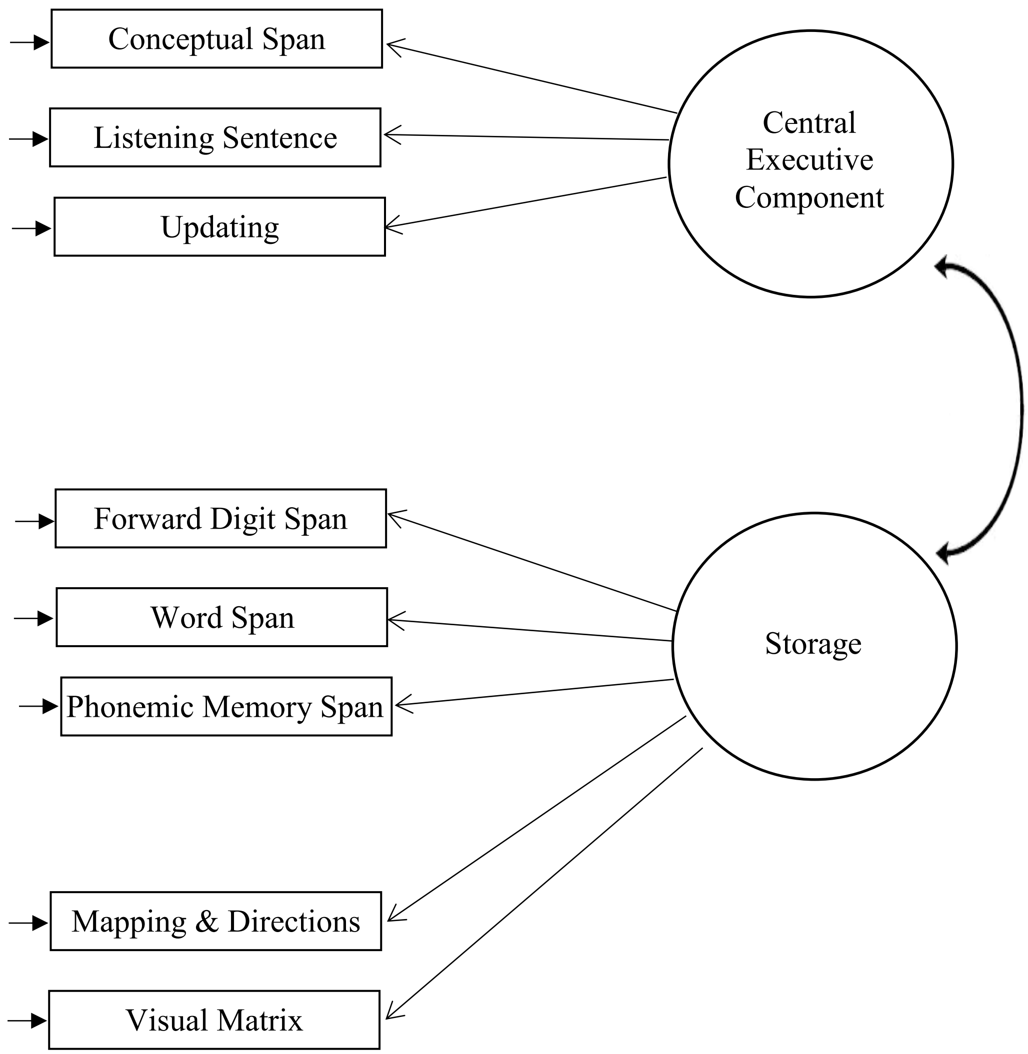

Figure 3. Baddeley's Model of Working Memory (Model 3).

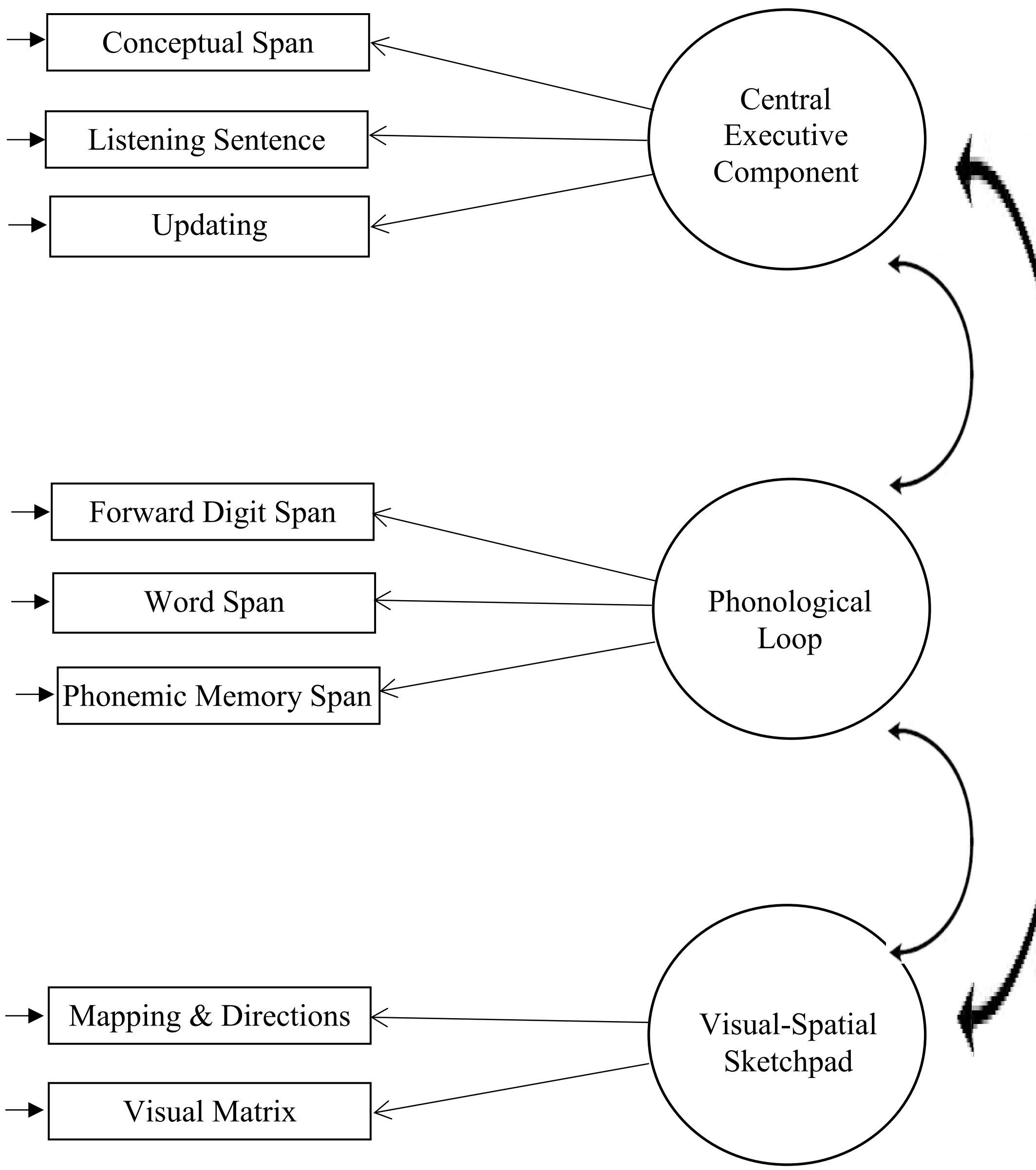

Table 1

Demographic and Psychometric Characteristics of Study Sample

	Age 6		Age 7		Age 8		Age 9		Age 10	
FedLunch	97%		94%		96%		97%		89%	
H/Lang	88%		92%		91%		90%		91%	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Age (in months)	78.17	5.54	90.95	3.90	103.42	5.21	114.96	3.59	127.37	5.10
Fluid Intelligence	57.59	20.10	61.94	23.18	59.12	23.20	62.80	22.53	59.07	23.22
Word Identification	106.26	12.36	102.93	15.87	98.31	15.20	100.56	17.66	99.49	17.33
Comprehension	92.78	15.85	93.14	15.67	89.11	13.08	90.38	13.66	90.92	13.15
E-Vocabulary	83.63	8.81	84.54	10.18	83.40	9.95	87.28	12.20	88.17	9.64
S-Vocabulary	86.85	18.06	80.72	15.48	78.80	15.43	75.22	15.54	77.48	19.68
Composite Language										
Male	92.24	12.83	90.42	10.65	90.89	10.52	91.57	11.35	90.80	11.05
Female	93.04	12.69	87.38	11.51	87.17	11.73	87.67	12.13	86.74	11.17

Note. FedLunch = Percentage of sample that participates in Federal Lunch Program; H/Lang = Percentage of sample that speak Spanish at Home; Word Identification = English Word Identification Subtest; Comprehension = English Reading Comprehension Subtest; E- = English; S = Spanish; Composite Language = Combined English and Spanish vocabulary scores.
Note. Except for Fluid intelligence (percentile scores) age (months), the remaining values are standard scores.

Table 2

Factor Models for Working Memory in English and Spanish as a Function of Age Group

One-Factor Model of Working Memory										
	English					Spanish				
	Age 6	Age 7	Age 8	Age 9	Age 10	Age 6	Age 7	Age 8	Age 9	Age 10
Chi-Square	59.40	32.84	49.07	32.00	25.26	81.35	20.44	46.83	40.05	41.54
Chi-Square *DF*	20	20	20	20	20	20	20	20	20	20
Pr > Chi-Square	< .001	0.04	< .001	0.04	0.19	< .001	0.43	< .001	< .001	< .001
Standardized RMSR (SRMSR)	0.10	0.08	0.08	0.09	0.07	0.14	0.06	0.08	0.09	0.08
RMSEA Estimate	0.12	0.07	0.10	0.09	0.05	0.15	0.01	0.10	0.10	0.09
RMSEA, Lower 90% Confidence	0.09	0.02	0.07	0.02	0.00	0.12	0.00	0.06	0.06	0.05
RMSEA, Upper 90% Confidence	0.16	0.11	0.14	0.14	0.10	0.19	0.08	0.13	0.15	0.14
Akaike Information Criterion	91.40	64.84	81.07	64.00	57.26	113.35	52.44	78.83	72.05	73.54
Bayesian Information Criterion	137.64	110.72	128.25	102.90	101.99	159.35	97.7	126.01	112.74	118.27
Comparative Fit Index	0.64	0.80	0.74	0.72	0.90	0.42	0.99	0.78	0.78	0.77

Two-Factor Model of Working Memory										
	English					Spanish				
	Age 6	Age 7	Age 8	Age 9	Age 10	Age 6	Age 7	Age 8	Age 9	Age 10
Chi-Square	54.57	27.45	34.04	28.45	21.27	78.33	20.39	46.34	39.90	40.45
Chi-Square *DF*	19	19	19	19	19	19	19	19	19	19
Pr > Chi-Square	< .001	0.09	0.02	0.08	0.32	< .001	0.37	< .001	< .001	< .001
Standardized RMSR (SRMSR)	0.10	0.07	0.07	0.08	0.06	0.14	0.06	0.08	0.09	0.08
RMSEA Estimate	0.12	0.06	0.08	0.08	0.03	0.16	0.02	0.10	0.11	0.10
RMSEA, Lower 90% Confidence	0.08	0.00	0.03	0.00	0.00	0.12	0.00	0.06	0.06	0.06
RMSEA, Upper 90% Confidence	0.16	0.10	0.12	0.13	0.09	0.19	0.08	0.14	0.16	0.14

(Table Continues)

Table 2 *(continued)*

Two-Factor Model of Working Memory *(continued)*	English					Spanish				
	Age 6	Age 7	Age 8	Age 9	Age 10	Age 6	Age 7	Age 8	Age 9	Age 10
Akaike Information Criterion	88.57	61.45	68.04	62.45	55.27	112.33	54.39	80.34	73.90	74.45
Bayesian Information Criterion	137.71	110.20	118.17	103.77	102.80	161.21	102.47	130.47	117.13	121.98
Comparative Fit Index	0.67	0.87	0.87	0.78	0.96	0.44	0.97	0.78	0.77	0.77

Three Factor Model of Working Memory	English					Spanish				
	Age 6	Age 7	Age 8	Age 9	Age 10	Age 6	Age 7	Age 8	Age 9	Age 10
Chi-Square	15.09	23.53	18.64	14.55	18.44	16.71	15.52	14.08	29.25	18.44
Chi-Square *DF*	17	17	17	17	17	17	17	17	17	17
Pr > Chi-Square	0.59	0.13	0.35	0.63	0.36	0.47	0.56	0.66	0.03	0.36
Standardized RMSR (SRMSR)	0.05	0.07	0.05	0.06	0.06	0.06	0.05	0.04	0.07	0.06
RMSEA Estimate	0.00	0.05	0.03	0.00	0.03	0.00	0.00	0.00	0.09	0.03
RMSEA, Lower 90% Confidence	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
RMSEA, Upper 90% Confidence	0.07	0.10	0.08	0.08	0.09	0.08	0.07	0.06	0.14	0.09
Akaike Information Criterion	53.09	61.53	56.64	52.55	56.44	54.71	53.52	52.08	67.25	56.44
Bayesian Information Criterion	108.01	116.01	112.67	98.73	109.56	109.34	107.26	108.10	115.57	109.56
Comparative Fit Index	1.00	0.90	0.99	1.00	0.97	1.00	1.00	1.00	0.86	0.97

Table 3

Working Memory z-Scores Partialed for Vocabulary as a Function of Gender, Memory Span, and Age Group

	Age 6		Age 7		Age 8		Age 9		Age 10	
	LSmean	*SE*	LSmean	*SE*	LSmean	*SE*	LSmean	*SE*	LSmean	*SE*
E-STM										
Male	-0.23	0.04	-0.02	0.04	0.18	0.05	0.30	0.05	0.43	0.05
Female	-0.16	0.04	0.03	0.04	0.18	0.04	0.27	0.05	0.43	0.04
S-STM										
Male	-0.13	0.04	-0.14	0.04	0.16	0.04	0.25	0.05	0.32	0.04
Female	-0.10	0.04	0.11	0.04	0.16	0.04	0.20	0.05	0.41	0.03
E-Exec										
Male	-0.18	0.03	0.01	0.04	0.15	0.05	0.26	0.06	0.42	0.07
Female	-0.16	0.03	0.01	0.03	0.14	0.04	0.33	0.06	0.43	0.06
S-Exec										
Male	-0.11	0.03	-0.07	0.03	0.05	0.04	0.23	0.05	0.42	0.06
Female	-0.10	0.02	0.02	0.03	0.16	0.03	0.30	0.05	0.41	0.05
Vis-Spatial										
Male	-0.14	0.03	0.04	0.03	0.15	0.05	0.60	0.07	0.80	0.08
Female	-0.14	0.03	-0.01	0.03	0.10	0.04	0.39	0.07	0.65	0.07
	df	*F*-ratio	*df*	*F*-ratio	*df*	*F*-ratio	*df*	*F*-ratio	*df*	*F*-ratio
Gender	5,124	0.34	5,123	4.01**	5,137	1.63	5,88	1.69	5,112	1.17
Span	4,125	2.49*	4,124	2.56*	4,138	2.39*	4,89	1.38	4,113	4.17**
Interaction	4,125	0.40	4,124	4.72**	4,138	1.92	4,89	2.07*	4,113	1.40

Note. LSmean = Least square means--means partialed for the influence of vocabulary.
*$p < .05$, **$p < .01$

Table 4

Correlations among Latent Variables by Age Group

		Age 6					Age 7				
		1	2	3	4	5	1	2	3	4	5
1	E-STM	1					1				
2	S-STM	0.49	1				**0.59**	1			
3	E-Exec	0.30	0.26	1			0.30	0.19	1		
4	S-Exec	0.37	**0.53**	0.45	1		0.28	0.40	0.28	1	
5	Vis-spatial	0.28	0.08	0.28	0.16	1	0.06	-0.02	0.15	0.03	1
		Age 8					Age 9				
		1	2	3	4	5	1	2	3	4	5
1	E-STM	1					1				
2	S-STM	**0.55**	1				**0.58**	1			
3	E-Exec	0.38	0.22	1			0.40	0.34	1		
4	S-Exec	0.37	**0.52**	0.37	1		0.41	**0.50**	**0.52**	1	
5	Vis-spatial	0.28	0.21	0.14	0.17	1	0.15	0.17	0.17	0.28	1
		Age 10									
		1	2	3	4	5					
1	E-STM	1									
2	S-STM	0.42	1								
3	E-Exec	0.41	0.49	1							
4	S-Exec	0.37	**0.53**	**0.58**	1						
5	Vis-spatial	0.19	0.05	0.05	0.13	1					

Note. E- = English; S- = Spanish; STM = Phonological Loop; Exec = Executive System; Vis-spatial = Visual-Spatial Working Memory.
Note. Parameters in bold yielded high effect sizes.

Table 5

Hierarchical Model Showing Parameter Estimates for First and Second Order Factor Model

	Age 6		Age 7		Age 8		Age 9		Age 10	
	S-Factor	G-factor	S-Factor	G-factor	S-Factor	G-factor	S-Factor	G-factor	S-Factor	G-factor
Phonological Loop										
E-FwrD	0.04	**0.54**	0.13	**0.61**	0.06	**0.68**	0.23	**0.60**	0.25	**0.42**
E-Words	0.38	**0.75**	0.22	**0.49**	0.08	**0.53**	0.41	**0.53**	0.22	**0.35**
E-Pseudo	0.10	**0.42**	0.05	0.24	0.31	**0.45**	0.40	**0.44**	0.54	**0.31**
S-FwrD	0.63	**0.78**	0.13	**0.75**	0.03	**0.78**	0.19	**0.73**	0.10	**0.55**
S-Words	0.16	**0.46**	0.80	**0.61**	0.12	**0.66**	0.06	**0.61**	0.10	**0.46**
S-Pseudo	0.14	**0.40**	0.07	**0.40**	0.95	**0.31**	0.35	**0.36**	0.48	**0.31**
Executive System										
E-Cspan	0.40	0.15	0.99	0.13	0.40	0.22	0.55	0.22	0.30	**0.41**
E-Lspan	0.47	0.10	0.34	0.04	0.57	0.10	0.47	**0.30**	0.51	**0.30**
E-Update	0.20	0.25	0.31	**0.50**	0.47	**0.37**	0.11	**0.35**	0.17	**0.70**
S-Cspan	0.35	0.26	0.27	0.14	0.29	**0.32**	0.44	0.05	0.46	**0.43**
S-Lspan	0.49	0.16	0.08	0.14	0.09	0.18	0.35	**0.39**	0.50	**0.33**
S-Update	0.09	**0.30**	0.03	**0.41**	0.21	**0.41**	0.09	**0.51**	0.15	**0.74**
Visual-Spatial Sketchpad										
Matrix	0.93	0.12	0.14	0.03	0.40	0.25	0.54	0.32	0.29	0.01
Mapping	0.57	0.08	.99	0.01	0.96	0.11	0.65	0.18	0.52	0.22

Note. S-Factor = Specific Factor; G-Factor = General Factor; E- = English; S- = Spanish; FwrD = Forward Digit Span task; Words = Word Span task; Pseudo = Pseudoword Span task; Cspan = Conceptual Span task; Lspan = Listening Sentence Span task; Update = Updating task; Matrix = Visual Matrix task; Mapping = Mapping & Directions task.
Note. Parameters in bold yielded moderate to large effect sizes.

Table 6

Factor Structure within Age Groups for English and Spanish Measures

Three-Factor Model					
	Age 6	Age 7	Age 8	Age 9	Age 10
Chi-Square	88.38	77.73	90.22	72.12	87.30
Chi-Square *df*	64	65	64	63	63
Pr > Chi-Square	0.02	0.13	0.02	0.20	0.02
Standardized RMSR (SRMSR)	0.07	0.07	0.06	0.07	0.06
RMSEA Estimate	0.05	0.04	0.05	0.04	0.06
RMSEA, Lower 90% Confidence	0.02	0.00	0.02	0.00	0.02
RMSEA, Upper 90% Confidence	0.08	0.07	0.08	0.08	0.08
Akaike Information Criterion	170.38	157.73	172.22	156.12	171.30
Bayesian Information Criterion	288.26	270.87	292.82	258.72	288.38
Comparative Fit Index	0.90	0.93	0.92	0.94	0.90
Four-Factor Model					
	Age 6	Age 7	Age 8	Age 9	Age 10
Chi-Square	80.89	81.86	89.17	72.25	86.52
Chi-Square *df*	63	66	63	64	63
Pr > Chi-Square	0.06	0.09	0.02	0.22	0.03
Standardized RMSR (SRMSR)	0.07	0.07	0.06	0.07	0.06
RMSEA Estimate	0.05	0.04	0.05	0.04	0.06
RMSEA, Lower 90% Confidence	0.00	0.00	0.02	0.00	0.02
RMSEA, Upper 90% Confidence	0.07	0.07	0.08	0.08	0.08
Akaike Information Criterion	164.89	159.86	173.17	154.25	170.52
Bayesian Information Criterion	285.65	270.16	296.72	254.40	287.59
Comparative Fit Index	0.92	0.92	0.92	0.95	0.90

Table 7

Multi-Sample Analysis of Measurement Invariance within Language Systems for Ages 6,7,8,9 and 10.

	English-Three Factor Model									
Configural Invariance										
Between Age Comparisons	6 vs. 7	6 vs.8	6 vs. 9	6 vs.10	7 vs. 8	7 vs. 9	7 vs. 10	8 vs. 9	8 vs. 10	9 vs. 10
Chi-Square	33.09	27.89	30.03	30.34	33.74	31.71	32.02	26.51	26.82	27.81
Chi-Square DF	34	34	34	34	34	34	34	34	34	34
Pr > Chi-Square	0.51	0.76	0.66	0.65	0.48	0.58	0.56	0.82	0.80	0.71
Standardized RMSR (SRMSR)	0.05	0.05	0.05	0.05	0.05	0.06	0.05	0.05	0.04	0.05
RMSEA Estimate	0	0	0	0	0	0	0	0	0	0
RMSEA Lower 90% Confidence	0	0	0	0	0	0	0	0	0	0
RMSEA Upper 90% Confidence	0.06	0.04	0.06	0.05	0.06	0.06	0.06	0.04	0.04	0.05
Comparative Fit Index	1	1	1	1	1	1	1	1	1	1
Pattern invariance										
Chi-Square	35.60	32.08	39.23	37.88	36.89	43.75	38.12	34.94	31.97	28.91
Chi-Square DF	39	39	39	39	39	39	39	39	39	39
Pr > Chi-Square	0.63	0.78	0.46	0.52	0.57	0.28	0.51	0.66	0.78	0.91
Standardized RMSR (SRMSR)	0.06	0.05	0.06	0.06	0.06	0.07	0.06	0.06	0.05	0.05
RMSEA Estimate	0	0	0.01	0	0	0.03	0	0	0	0
RMSEA Lower 90% Confidence	0	0	0	0	0	0	0	0	0	0
RMSEA Upper 90% Confidence	0.05	0.04	0.07	0.06	0.05	0.08	0.06	0.05	0.04	0.03
Comparative Fit Index	1	1	1	1	1	0.96	1	1	1	1
$\Delta\chi^2$	2.51	4.18	9.02	7.54	3.15	12.04*	6.10	8.43	5.15	1.10
	Spanish-Three Factor Model									
Configural Invariance										
Chi-Square	24.29	22.99	27.27	35.69	25.37	38.01	38.11	36.71	36.73	49.41
Chi-Square DF	34	34	34	34	34	34	34	34	34	34
Pr > Chi-Square	0.89	0.92	0.92	0.39	0.86	0.29	0.29	0.34	0.34	0.04
Standardized RMSR (SRMSR)	0.05	0.04	0.05	0.05	0.05	0.06	0.06	0.05	0.05	0.07
RMSEA Estimate	0	0	0	0.02	0	0.03	0.03	0.03	0.02	0.07
RMSEA Lower 90% Confidence	0	0	0	0	0	0	0	0	0	0.01

RMSEA Upper 90% Confidence	0.03	0.02	0.02	0.07	0.04	0.08	0.08	0.07	0.07	0.1
Comparative Fit Index	1	1	1	0.99	1	0.96	0.97	0.98	0.99	0.90
Pattern invariance										
Chi-Square	34.13	27.27	44.95	44.36	33.69	57.15	50.32	50.08	48.18	63.17
Chi-Square DF	39	39	39	39	39	39	39	39	39	39
Pr > Chi-Square	0.69	0.92	0.24	0.26	0.71	0.03	0.11	0.11	0.15	0.01
Standardized RMSR (SRMSR)	0.06	0.05	0.07	0.06	0.06	0.09	0.07	0.08	0.07	0.09
RMSEA Estimate	0	0	0.04	0.03	0	0.07	0.05	0.05	0.04	0.08
RMSEA Lower 90% Confidence	0	0	0	0	0	0.02	0	0	0	0.04
RMSEA Upper 90% Confidence	0.05	0.02	0.08	0.07	0.05	0.1	0.08	0.09	0.08	0.11
Comparative Fit Index	1	1	0.97	0.97	1	0.84	0.91	0.93	0.95	0.85
$\Delta\chi^2$	9.84	4.28	*17.68	**8.67**	8.32	*13.14	12.21*	*13.37	11.45*	*13.76

Note. * p < .05, $\Delta\chi^2 > 11.07$, pattern invariance (equivalent factor loadings) between age groups not supported. Bold did not meet Cheung & Rensvold (2002) criterion for accepting pattern invariance model .

Table 8

Multi-Sample Analysis of Measurement Invariance Across Language Systems for Ages 6,7,8,9, and 10.

Between Age Comparisons	6 vs.7	6 vs. 8	6 vs. 9	6 vs. 10	7 vs 8	7 vs 9	7 vs 10	8 vs 9	8 vs 10	9 vs.10
Configural Invariance										
Chi-Square	124.93	107.77	108.19	113.6	107.8	108.22	121.5	85.89	91.3	96.89
Chi-Square DF	92	92	92	92	92	92	92	92	92	92
Pr > Chi-Square	0.01	0.12	0.12	0.06	0.12	0.12	0.02	0.66	0.5	0.34
Standardized RMSR (SRMSR)	0.06	0.05	0.06	0.06	0.06	0.07	0.07	0.06	0.05	0.06
RMSEA Estimate	0.05	0.04	0.04	0.04	0.04	0.04	0.05	0	0	0.02
RMSEA Lower 90% Confidence	0.03	0	0	0	0	0	0.02	0	0	0
RMSEA Upper 90% Confidence	0.08	0.06	0.07	0.07	0.06	0.07	0.07	0.04	0.05	0.06
Comparative Fit Index	0.91	0.96	0.95	0.95	0.96	0.95	0.93	1.0	1.0	0.99
Pattern invariance										
Chi-Square	136.51	117.86	115.63	119.23	113.32	121.94	128.02	103.28	108.9	105.96
Chi-Square DF	99	99	99	99	99	99	99	99	99	99
Pr > Chi-Square	0.01	0.1	0.12	0.08	0.15	0.06	0.03	0.36	0.23	0.3
Standardized RMSR (SRMSR)	0.07	0.06	0.07	0.06	0.06	0.08	0.07	0.07	0.06	0.07
RMSEA Estimate	0.05	0.04	0.04	0.04	0.03	0.05	0.05	0.02	0.03	0.03
RMSEA Lower 90% Confidence	0.03	0	0	0	0	0	0.02	0	0	0
RMSEA Upper 90% Confidence	0.08	0.06	0.07	0.06	0.06	0.07	0.07	0.05	0.06	0.06
Comparative Fit Index	0.89	0.95	0.95	0.95	0.96	0.93	0.93	0.99	0.98	0.98
$\Delta\chi^2$	**11.57**	10.09	7.44	5.63	5.52	**13.72**	6.52	17.39*	17.60*	9.07

Note. * $p < .05$, $\Delta\chi^2 > 14.07$, pattern invariance (equivalent factor loadings) between age groups not supported. Bold: hypothesis did not meet Cheung & Rensvold (2002) criterion for accepting the pattern invariance model .

Appendix A

Sample Size, Means, and Standard Deviations of English and Spanish Measures and Correlation Matrix as a Function of Age

Age 6 Note: Means and Standard Deviations from 133 students (61 males and 72 females).

Variable	Name	Mean	SD	KR_{20}	1	2	3	4	5	6	7	8
1	E-Digits	3.18	0.83	0.92	1							
2	E-Word	2.13	0.77	0.83	0.4	1						
3	E-Pseudo	1.04	0.63	0.86	0.13	0.33	1					
4	E-Concept	2.16	1.8	0.95	0.11	0.14	0.14	1				
5	E-Listen	1.44	0.27	0.57	0.1	0.01	-0.04	0.09	1			
6	E-update	1.49	1.62	0.95	0.27	0.22	0.09	0.14	0.19	1		
7	Visual Matrix	8.15	5.43	0.94	0.18	0.07	0.12	0.23	0.14	0.15	1	
8	Map/ Directions	1.77	1.25	0.86	0.08	0.07	0.05	0.11	0.15	0.05	0.6	1
9	S-Digits	3.16	0.64	0.62	0.43	0.34	0.27	0.05	0.1	0.25	0.05	-0.01
10	S-word	1.57	0.78	0.91	0.14	0.29	0.13	0.09	0.07	0.07	0.05	0.02
11	S-pseudo	1.11	0.61	0.85	0.21	0.32	0.36	0.11	0.09	0.04	0.11	0.09
12	S-Concept	1.42	1.57	0.91	0.11	0.19	0.01	0.26	0.12	0.08	0.18	0.17
13	S-Listenin	0.48	0.68	0.98	-0.07	0.05	-0.01	0.2	0.25	0.07	0.09	0.03
14	S-update	0.96	1.63	0.94	0.12	0.21	-0.09	-0.01	0.09	0.16	0.08	-0.01

Age 7 Note: Means and Standard Deviations from 130 students (62 males and 68 females).

	name	Mean	SD	KR_{20}	1	2	3	4	5	6	7	8
1	E-Digits	3.66	0.85	0.71	1							
2	E-Word	2.27	0.68	0.79	0.31	1						
3	E-Pseudo	1.16	0.6	0.86	0.22	0.12	1					
4	E-Concept	3.01	2.37	0.91	0.06	0.2	0.14	1				
5	E-Listen	1.61	0.35	0.57	-0.03	0.07	-0.02	0.3	1			
6	E-update	2.23	1.69	0.89	0.32	0.28	0.17	0.32	0.11	1		
7	Visual Matrix	11.37	6.13	0.92	0.01	0.01	0.01	0.15	0.06	0.14	1	
8	Map/ Directions	2.62	1.6	0.85	-0.06	-0.12	-0.02	-0.05	-0.11	-0.05	0.19	1
9	S-Digits	3.34	0.8	0.83	0.49	0.37	0.11	0.08	-0.07	0.33	-0.01	0.05

10	S-word	1.84	0.77	0.93	0.23	0.43	0.15	0.08	-0.06	0.34	-0.01	0.01
11	S-pseudo	1.21	0.72	0.91	0.09	0.13	0.18	-0.01	0.06	0.18	-0.02	0.11
12	S-Concept	2	1.58	0.89	0.08	0.01	0.1	0.28	0.11	0.09	0.01	0.05
13	S-Listen	0.68	0.92	0.91	-0.15	0.02	-0.12	0.04	0.12	-0.21	-0.01	0.01
14	S-update	1.33	1.76	0.94	0.19	0.2	-0.08	0.07	0.03	0.25	-0.04	0.14
Age 8		Note: Means and Standard Deviations from 141 students (61 males and 80 females).										
	name	Mean	SD	KR_{20}	1	2	3	4	5	6	7	8
1	E-Digits	3.83	0.89	0.67	1							
2	E-Word	2.56	0.77	0.85	0.39	1						
3	E-Pseudo	1.4	0.75	0.91	0.29	0.28	1					
4	E-Concept	4.47	3.48	0.92	0.22	0.15	0.12	1				
5	E-Listen	1.74	0.44	0.66	0.09	0.08	0.09	0.26	1			
6	E-update	2.73	2.05	0.92	0.3	0.19	0.2	0.29	0.21	1		
7	Visual Matrix	13.49	7.06	0.93	0.14	0.15	0.13	0.12	0.08	0.13	1	
8	Map/ Directions	3.23	3.22	0.92	0.06	0.14	0.26	0.13	0.08	0.05	0.38	1
9	S-Digits	3.56	0.8	0.83	0.51	0.43	0.33	0.11	0.08	0.21	0.04	-0.08
10	S-word	1.96	0.84	0.93	0.4	0.19	0.33	0.27	0.06	0.15	0.16	0.02
11	S-pseudo	1.35	0.68	0.91	0.12	0.23	0.39	0.04	0.03	0.01	0.11	0.09
12	S-Concept	2.33	2.16	0.89	0.21	0.11	0.14	0.26	0.2	0.13	0.13	-0.04
13	S-Listenin	0.92	0.89	0.91	0.05	0.08	0.12	0.05	0.18	0.1	0.05	0.01
14	S-update	2.12	2.31	0.94	0.16	0.31	0.18	0.1	0.19	0.24	0.28	0.04
Age 9		Note: Means and Standard Deviations from 94 students (46 males and 48 females).										
	name	Mean	SD	KR_{20}	1	2	3	4	5	6	7	8
1	E-Digits	4	0.93	0.73	1							
2	E-Word	2.78	0.8	0.91	0.45	1						
3	E-Pseudo	1.52	0.71	0.9	0.21	0.17	1					
4	E-Concept	4.01	2.49	0.93	0.11	0.05	0.08	1				
5	E-Listen	1.92	0.46	0.63	0.22	0.23	0.11	0.34	1			
6	E-update	3.23	2.55	0.95	0.18	0.09	0.14	0.17	0.12	1		
7	Visual Matrix	17.74	8.34	0.93	0.06	0.08	0.04	-0.01	0.17	0.11	1	
8	Map/ Directions	7.21	6.38	0.97	0.17	-0.01	-0.11	0.03	0.16	0.18	0.36	1

9	S-Digits	3.75	0.83	0.86	0.54	0.49	0.29	0.18	0.2	0.24	0.15	0.11
10	S-word	2	0.82	0.94	0.32	0.35	0.29	0.19	0.13	0.26	0.08	0.14
11	S-pseudo	1.54	0.72	0.89	0.23	0.04	0.28	0.28	0.07	0.04	-0.05	0.21
12	S-Concept	2.85	1.53	0.84	-0.09	0	0.05	0.29	0.16	-0.02	0.13	0.01
13	S-Listen	1.45	1.51	0.93	0.2	0.2	0.08	0.17	0.35	0.25	0.28	0.24
14	S-update	2.72	2.76	0.97	0.29	0.2	0.28	0.13	0.18	0.33	0.12	0.11
Age 10		Note: Means and Standard Deviations from 121 students (52 males and 69 females).										
	Name	Mean	SD	KR_{20}	1	2	3	4	5	6	7	8
1	E-Digits	4.2	0.87	0.56	1							
2	E-Word	2.95	0.73	0.86	0.26	1						
3	E-Pseudo	1.75	0.76	0.91	0.2	0.21	1					
4	E-Concept	5.27	4.21	0.91	0.12	0.1	0.17	1				
5	E-Listen	1.99	0.55	0.71	0.02	0.18	0.13	0.31	1			
6	E-update	3.72	2.52	0.95	0.31	0.23	0.19	0.26	0.15	1		
7	Visual Matrix	20.49	8.57	0.96	0.05	-0.05	0	0.02	0.02	-0.1	1	
8	Map/ Directions	9.73	8.34	0.95	0.12	0.15	0.15	0.06	0	0.14	0.19	1
9	S-Digits	3.86	0.74	0.8	0.45	0.18	0.26	0.17	0.25	0.34	0.04	-0.06
10	S-word	2.33	0.83	0.76	0.19	0.25	0.12	0.33	0.16	0.28	-0.03	-0.03
11	S-pseudo	1.8	0.76	0.91	0.25	0.24	0.39	0.05	0.06	0.32	-0.01	0.2
12	S-Concept	3.77	2.8	0.88	0.13	0.14	0.11	0.26	0.38	0.21	-0.02	-0.02
13	S-Listen	1.92	2.27	0.94	0.12	0.02	0.24	0.19	0.32	0.11	-0.1	0.2
14	S-update	3.05	2.6	0.96	0.28	0.22	0.28	0.26	0.21	0.56	0.01	0.21

Note. E-English, S-Spanish, Pseudo=pseudo or nonword span; Concept=Conceptual span, Listen=Listening span, Map/Directions=Mapping and directions
KR_{20}=Kuder-Richardson reliabilities