Growth in Math Computation among Monolingual and English Language Learners:

Does the Executive System have a Role?

Growth in Math Computation among Monolingual and English Language Learners: Does

Executive Processing Play a Role?

H. Lee Swanson

Jennifer Kong

Stefania D. Petcu

University of New Mexico

H. Lee Swanson, Ph.D., Educational Psychology, University of California-Riverside (Lee.Swanson@ucr.edu) and University of New Mexico (HLswanson@unm.edu). Jennifer Kong, Ph.D., Special Education, University of California-Riverside and post-doctoral University of New Mexico (JenniferKong@unm.edu) and Stefania D. Petcu, Ph.D., Office of Research and Community Engagement, University of New Mexico (spetcu@unm.edu) .

This research is based on two four-year longitudinal studies funded by the U.S. Department of Education, Cognition and Student Learning (R305H020055; R324A090092), Institute of Education Sciences (IES). This earlier foundational work is being extended under a National Science Foundation (NSF) grant, Division of Research on Learning (award number 1660828) Sciences awarded to the first author. This study does not necessarily reflect the views of the U.S. Department of Education, IES, NSF, or the participating school districts. Special appreciation is given to Margaret Beebe-Frankenberger, Cathy Lussier, Bev Hedin, Lelani Saez, Olga Jerman, Xinhua Zheng, Melina, Melgarejo, Joseph Rios, Elizabeth Arellano, Nicole Garcia, Alfredo Aviles, Steve Gómez, Paula Aisemberg, Valerie Perry, Loren Albeg, Dennis Sisco-Taylor, Wenson Fung, Georgia Doukas, Rebecca Greg, Kelly Rosston, Crystal Howard, Mike Gerber, Michael Orosco, and School District Laison and Consultant: Erin Bostick Mason for data collection and/or analysis.

Correspondence concerning this article should be addressed to H. Lee Swanson, Educational Psychology, College of Education, University of New Mexico, and Albuquerque, NM 87131.

E-mail:HLswanson@unm.edu

Abstract

This cohort-sequential study explored the components of working memory that underlie math calculation in elementary school children who are monolingual (English) or English language learners (ELLs) whose first language is Spanish. To this end, children (*N* = 789) in grades 1, 2, and 3 at wave 1 were administered a battery of math, vocabulary, reading and cognitive (short-term memory [STM], working memory [WM], rapid naming, and inhibition) measures. The battery of tests was administered again one year and two years later to the same participants. Three important findings emerged. First, along with naming speed, the results suggest that growth in the executive component of WM was significantly related to growth in calculation performance. Second, performance on measures of reading, fluid intelligence, naming speed and executive processes in wave 1 were significantly related to wave 3 math calculation performance. Finally, the full latent growth model showed that monolingual and ELL children were statistically comparable in computation at wave 3. Thus, strong support was found for the notion that the executive component of WM was related to math computation but weak support for the notion ELL children experienced a math achievement gap.

Keywords: Math Disabilities, English Language Learner (ELL), Working Memory, Cognition, Math Computation

Growth in Math Computation among Monolingual and English Language Learners:

Does the Executive System have a Role?

Math computation is a fundamental skill that underlies later math skills, such as algebra and math problem solving (e.g., Cowan, Donlan, Shepherd, Cole-Fletcher, Saxton, & Hurry, 2011; Fuchs et al, 2016; Geary, Nicholas & Sun, 2017; Powell, Kearns & Driver, 2016). Further, computational knowledge is a fundamental skill used to identify children with serious math difficulties in monolingual children (e.g., Geary, Hoard, Nugent & Bailey, 2012) as well English language learners (ELLs, e.g., Vukovic & Lesaux, 2013)[1]. Children designated as ELL with Spanish as a first language in the United States have consistently yielded low mathematics scores when compared to other ELL groups on national assessments over several decades (e.g., Hemphill & Vanneman, 2011; NAEP, National Assessment of Education Progress, 2004, 2015, 2017; also see Roberts & Bryant, 2011). No doubt, there are long-term implications related to this achievement gap, such as detrimental effects on high school performance (e.g., drop rates) as well as later employment (e.g., Grégoire & Desoete, 2009; Polk, 2016), and therefore further research is necessary to determine processes that may underlie low math computation performance in order to develop effective interventions.

Although some of the difficulties in math computation experienced by monolingual and ELL children have been attributed to oral vocabulary and/or reading skill (e.g., Gorman, 2012; Kempert, Saalbach, & Hardy, 2011; Macizo, Herra, Roman, & Martin, 2011; Martiniello, 2009), other processes besides language and reading may play a critical role in understanding such children's math performance. There is recent evidence to suggest that working memory (WM) plays a significant role in math performance for monolingual children who suffer from serious math difficulties (e.g., Cowan et al., 2011; Peng Namkung, Barnes & Sun, 2016; Klesczewski et al., 2015; Swanson & Beebe-Frankenberger, 2004) as well as ELL children whose first language is Spanish (e.g., ; Harvey & Miller, 2017; Vukovic & Lesaux, 2013; Swanson, Kong, & Petcu, 2018a). Working memory is defined as a limited capacity system of temporary stores that includes functions related to the preservation of information while simultaneously processing other information and attention control related to these functions (e.g., Baddeley, 2012).

The executive component of WM (tasks that involve simultaneous storage and processing) has been associated with math difficulties in both monolingual and ELL children (e.g., Blairm, Ursache, Greenberg, & Vernon-Feagan, 2015; Bull & Scerif, 2001; Cai, Georgiou, Wen, & Das, 2016; David, 2012; Menon, 2016; Swanson & Beebe-Frankenberger, 2004). One framework to capture executive processing as they apply to monolingual and ELL children is Baddeley's multicomponent WM model (Baddeley, 2012; Baddeley & Logie, 1999; also see Swanson, Kudo, & Van Horn, 2018b, for discussion). 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 of limited duration 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). 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 monolingual children (e.g., Gathercole, Pickering, Ambridge, & Wearing, 2004; Gray et al., 2017) and ELL children (e.g., Swansonet al., 2018b).

Bilingual children, however, are viewed as experiencing some advantages in executive processing when compared to monolingual children (e.g., Bialystok, 2011; Blom, Kuntay, Messer, Verhagen & Leseman, 2014). Clearly, ELL children (emerging bilinguals) vary in their mastery level of each language which in turn may play a role in their executive processing performance (Cummins, 1979; Kempert et al, 2011; Lonigan, Allan, Goodrick, Farrington & Phillips, 2017; Rosselli, Aridila, Lalwani & Velez-Uribe, 2016). Thus, the advantages that executive processing may have on achievement that has been attributed to bilingual children may not apply to some ELL children. For example, although some studies demonstrate a bilingual benefit in WM (e.g., Morales et al., 2013; Kudo & Swanson, 2012), not all studies find such an effect (e.g., de Abreu, Pascle, & Engle, 2011; de Bruin, Barbara & Della Sala, 2015; Namazi & Thordardotir, 2010). Thus, whether the executive component of WM plays a critical role in growth in math computation performance among both ELL and monolingual children is unclear.

In addition to the lack of clarity related to advantages of bilingualism on WM performance, there is also some confusion as to whether the executive component of WM or bilingualism plays a significant major role in math computation performance. For example, Fuchs and colleagues (2006) studied the cognitive correlates of arithmetic computation and arithmetic word problems in third graders. The significant contribution of WM to math performance was partialed out in the regression model when reading measures were entered into the analysis. Likewise, Peng et al. (2016) compared the development of 176 monolingual children identified with reading and math difficulties beginning in the first grade. Data collection included measures of WM, language, nonverbal reasoning, processing speed, socioeconomic status, and gender as well as calculation ability at four-time points. Their results suggested that processing speed and decoding skills (phonological processing) played a major role in the calculation performance in the first grade whereas numerical competence (object counting, line estimation) and processing speed explained significant variance in calculation performance at the end of third grade. More critically, the study also found that WM, language, and nonverbal reasoning were “not” significant predictors of growth in calculation performance. These findings are consistent with the earlier studies by Fuchs et al. (2005) which did not find cognitive skills (i.e., WM) predicted calculation performance among third-graders when measures of vocabulary and reading subskills were entered into the analysis.

Likewise, the link between bilingualism and computation is unclear. For example, a longitudinal study (Vukovic & Lesaux, 2013) focused on the role of language on an array of math skills (arithmetic, analysis/ probability, prealgebra, geometry) among 75 English speakers and 92 language minority learners followed from first to fourth grade. As expected, English speakers outperformed language minority readers at various time points. However, the results suggested a strong longitudinal relationship between language and higher order components of math (pre-algebra, probability thinking, and geometry), but not with arithmetic calculation. Their results suggested that language ability was not directly involved in how to manipulate qualities and execute algorithms related to arithmetic computation. Further, the effects of language (performance on the Peabody Picture Vocabulary Test) on gains in mathematical cognition did not differ between the groups. Although their findings contrasted with the assumption that there is a direct link between language and arithmetic performance, their results raised the question as to whether bilingual proficiency (i.e., receptive and expressive vocabulary proficiency in two or more languages) can be linked to proficiency in arithmetic computation.

The purpose of this study was to determine whether growth in math performance is related to growth in WM in monolingual and bilingual children. Because some studies have suggested that WM does *not* contribute unique variance to math computation, we systematically compared the various competing models that are assumed to partial out the influence of WM on computation performance. Three models are considered as an explanation of the role of WM in children's growth in math computation: one focuses on the child's skills in other areas such as reading and fluid intelligence, another focuses on the phonological storage or short-term memory (STM), and the third focuses on the inhibition component of WM. The models are not necessarily exclusive of one another (each process can contribute important variance to computation to some degree), but suggest that some processes are more important than others (see Swanson, 2015; Swanson & Fung, 2016; Swanson, Jerman, & Zheng, 2008; Swasnon, Orosco, Lussier, 2012; 2015 for review of each model).

Crystallized and Fluid Intelligence

The first model assumes that math computation performance is critically related to measures of fluid and crystallized intelligence and performance on these measures underlies differences between children who vary in computation skills. In addition, the model suggests WM plays a nonsignificant role in predictions of math computation when measures of fluid and crystallized intelligence are entered into a regression analysis. Fluid intelligence represents reasoning and the ability to solve novel problems, whereas crystallized intelligence represents acculturation, schooling and language development (e.g., Horn & Noll, 1997). Fluid intelligence is assessed on measures that tap reasoning, thinking, or the ability to acquire new knowledge, whereas crystallized intelligence is assessed on measures that tap what has been learned in a particular domain (e.g., reading and math; Carroll, 1993). Strong verbal abilities, such as reading, are known to load on measures of crystallized intelligence (Carroll, 1993), whereas tasks that tap nonverbal reasoning (e.g., Raven Colored Progressive Matrices Test) load on measures of fluid intelligence. Both of these forms of intelligence have been found to mitigate the influence of WM on computation performance (e.g., Fuchs et al., 2005; Peng et al., 2016).

Phonological Storage

The second model assumes that although measures related to fluid and crystallized intelligence may play an important role in the development of computation skills, fundamental differences in the development of the storage component of WM underlies computation performance. Indirect evidence for this assumption comes from studies linking the phonological system to a number of academic processes (e.g., mental computation, reading, vocabulary) during the early childhood years (e.g., Baddeley et al., 1998), and therefore high order activities (e.g., executive processing activities attributed to the WM system) of children are constrained by inefficiencies in the phonological STM storage (i.e., phonological loop) system. For example, several studies have linked phonological STM to vocabulary development in monolinguals (e.g., Baddeley, Gathercole, & Papagno, 1998) and ELL children (e.g., Farnia & Geva, 2016). Recent meta-analyses (e.g., Peng, Wang, & Namkung, 2018) also suggest that compared to typically developing individuals, children's math difficulties are related to such processes as phonological processing, naming speed, and short-term memory (e.g., also see Geary, Hoard, Byrd-Craven, Nugent & Numtee, 2007).

Executive Process Model

In contrast to the above models, the third model views growth in the executive component of WM (i.e. controlled attention), independent of the storage component, as playing a major role in children's growth in computation. However, one possible source of this independent contribution has been related to inhibition (e.g., Bull & Scerif, 2001; Bonifacci et al., 2011; Lonigan et al, 2017). Because inhibition has been attributed to WM (e.g., Engle, 2002; Friedman, Haberstick, Willcutt, Miyake, Young, & Hewitt, 2007), individual differences related to inhibition may play an important role in children's computation performance. Inhibition may also account for computation differences found among some ELL children. A number of studies have shown proficiency in L1 and L2 positively influence intentional control, cognitive flexibility and overall executive processing (e.g., Bialystok, 2011). A key mechanism in this process is the inhibition of the competing language system (Bialystok, 2011; Bialystok & Feng, 2009; Bonifacci, Giombini, Bellocchi, & Contento; 2011). Thus, difficulties in inhibition (e.g., activating the targeted language while inhibiting the second language) may account for the incidence of math difficulties in some ELL children.

There are clear expectations for this third model. The influence of WM on measures of computation in children follows automatically with improvements on measures of inhibition. To assess this possibility, we administered a random generation task. During random generation, the central executive acts as a rate-limited filtering device that filters out habitual responses (such as "1, 2, 3, 4" or "a, b, c, d") so that random responses can be generated (Baddeley, 2007; Cooper, 2016; Jahanshahi, Saleem, Ho, Dirnberger, & Fuller, 2006; Towse & Cheshire, 2007). Thus, because it is assumed that the executive component of WM is related to inhibition, entering inhibition measures into a regression model would eliminated the influence of the executive component of WM on computation performance.

Purpose of Study

Clearly, the aforementioned models overlap in the role they play in accounting for the relationship between growth in WM and growth in calculation growth among ELL and monolingual children. However, the purpose of this study was to determine if growth in math performance is related to growth in WM in monolingual and ELL children who vary in math computation ability. To this end, this study addresses two major questions.

Question 1: Is growth in math computation related to growth in the executive component of working memory, independent of measures of fluid/crystallized intelligence, STM, naming speed and inhibition?

On the basis of the aforementioned models, we considered three possibilities: (a) growth in the relationship between WM and calculation skill is superseded by measures of fluid and crystallized intelligence; (b) growth in WM and calculation skill is superseded by the storage component of WM (i.e., phonological loop); or (c) growth in WM and calculation skill is superseded by growth in inhibition. The approach used in this study to assess whether WM plays a major role in accounting for math differences in performance was to remove statistically that system's influence from the analysis. In this study, the influence of the fluid/crystallized intelligence, phonological storage (e.g., STM, naming speed) and inhibition was partialed via a hierarchical regression analysis between calculation and WM. We reasoned that if WM and calculation are primarily a function of measures of fluid and crystallized intelligence, the phonological storage system and/or inhibition, then the predictions of computation performance by performance on WM measures should be nonsignificant when the aforementioned measures are entered into the regression analysis. However, if growth in the central executive system of WM (executive processing) is uniquely related to growth in computation, then the correlations between these two variables will remain significant when the influence of fluid /crystallized, phonological storage and inhibition measures are partialed-out in the analysis.

2: Are there math computation advantages for ELL children who are relatively proficient bilinguals than less proficient ELL or monolingual children?

The extant literature suggests bilingualism is associated with enhanced cognitive effects, most evident in executive processing tasks (e.g., Bialystok, 2011; Bialystok & Feng, 2009; Bonifacci et al., 2011). Because executive processing is associated with math computation (e.g., David, 2012; Menon, 2016), and the executive component of WM is related to growth in math (e.g., Swanson et al., 2008), one might expect that relatively bilingually proficientELL children may have some advantage in computation growth when compared to monolingual and/or less proficient bilingual children. That is, joint language activation among ELL children who are relatively proficient in both languages may facilitate the updating of computational knowledge not found in monolingual and or less bilingually proficient ELL children. Because a "math achievement gap" has been a common finding when comparing monolingual and ELL samples in the U.S, it was important to determine whether this math achievement gap over time is related to bilingual proficiency or is merely an artifact of not controlling for some aforementioned variables (e.g., reading, fluid intelligence, executive processes). Therefore, it was important to examine whether there are some disadvantages in math computation performance that occur in ELL children when compared to monolingual children are related to some of the variables found in previous studies that have moderated (or eliminated) the significant relationship between WM and computation growth. For example, if differences in computation that emerge between ELL and monolingual children are eliminated by entering executive processing (executive component of WM) measures into the analysis, then these findings would suggest that some of the residual differences between the two group are related to executive processing.

Method

Participants

Eight hundred and forty-one (*N*=841) children in grades 1 (*n* = 288), 2 (*n* = 240) and 3 (*n* = 313) from six large school districts in the southwest United States participated in the first wave of this federally funded longitudinal study (Swanson et al., 2008; Swanson et al., 2015). The sample included monolingual (N=354) and ELL children (*N*=487). The sample included 431 females and 410 males. Classroom instruction in math, as well as reading, was in English. Fifty (50%) of the monolingual sample and ninety-six (96%) percent of the ELL sample participated in a federally supported free and/or reduced lunch program.

For the final testing wave (2 years later), the remaining sample included 716 children in grades 3 (*n* = 237), 4 (*n* = 204) and 5 (*n* = 268). Seven children repeated the second grade. The attrition of children who dropped out of the study was due to moving out of the school district. Because of the attrition, comparisons were made between children retained (*N*=716) and those not retained (*N*=125) on measures of achievement, chronological age, and gender representation. A MANOVA comparing retained and non-retained children was computed on math (computation), reading (word identification, passage comprehension), and fluid intelligence standard normed referenced measures at wave 1. No significant differences on norm-referenced achievement measures occurred between the retained and non-retained sample, Wilks' λ =.99, F (4, 801) = 1.40, p = .23. No significant difference occurred at wave 1 between children retained and not retained) as a function of gender, χ^2 (1, N = 841) = .08, p = .77, ethnicity χ^2 (5, N =841) = 7.12, p = .21, or chronological age, F (1, 839) =.71, p =.40. Thus, we assume that missingness related to the demographic variables was at random.

Bilingual proficiency. There is a lack of consensus on the definition of ELL in terms of capturing bilingual proficiency (e.g., whether to use expressive and/or receptive language, the frequency of English spoken at home.). All ELL children in this study were identified by their school district based on their low performance on the California English Language Development Test (CELDT) that measures English proficiency.[2] However, our individual testing of ELL children on normed referenced English and Spanish vocabulary measures yielded variations in children's proficiency in the two language systems (English and Spanish). Some ELL children yield comparable proficiency (based on norm-referenced scores) in both languages (proficient bilingualism) and some ELL children were less proficient in one language or both languages (less proficient bilinguals). The literature varies in methodology for determining proficient and less proficient bilingualism (e.g., Peña, Bedore, & Kester, 2016; Rosselli, Lalwani, & Vélez-Uribe, 2016), with some studies suggesting a focus on proficiency within each language system that in turn yields a total score and/or for focusing on conceptual proficiency (tests that allow for responses in either language) that yields highest score in the preferred language (see Peña et al., 2016; for a comprehensive review). Conceptual scoring was used in the current study because it combines children's vocabulary knowledge across languages. According to Peña et al (2016), the advantage of this approach is that it gives credit to word knowledge across two languages (capturing breadth of vocabulary) and provides a better comparison (in contrast separate proficiency in each language) with single language scores of monolinguals. The present study administered normed referenced measures to ELL children to tap conceptual language (*Expressive One-Word Picture Vocabulary Test-Spanish Bilingual Edition, 2001)* as well as normed referenced measures that specifically tapped language proficiency within each language system (i.e., Peabody Picture Vocabulary test [PPVT-4, Dunn & Dunn, 2007]; Test de Vocabulario en Imagenes, TVIP [Dunn, Lugo, Padilla & Dunn, 1986]) and the Expressive One-Word Picture Vocabulary Test–Spanish Bilingual Edition (EOWPVT-SBE) (Brownell, 2001). (These measures are discussed below).

For the present study, and similar to other studies (e.g., Rosselli et al., 2016), we relied on the median score across all the vocabulary language measures to separate the ELL children into proficient and less proficient bilinguals. Those participants with a total score in both languages above the median were regarded as "proficient bilinguals" and those below the median as "less proficient bilinguals". Table 1 shows the norm-referenced scores among the three groups: monolingual, proficient bilingual, less proficient bilinguals. As shown, median normed scores for the relatively bilingually proficient ELLchildren were in the normal range (> 85 standard score), whereas less bilingually proficientELL children yield a median language score in the below average range (< 85 standard score). Median scores rather than mean scores were used to divide the groups to control for outliers as well as the potential skewed distribution in the data (i.e. as shown in Table 1, expressive Spanish scores were low for both proficient and less proficient bilinguals). According to Cohen's (1988) criteria, the median scores between the two groups of ELL children was large (ES=2.43). The median score across the four tests for the total sample was 79.95 (SD=13.05). Although the median score reflected a combination of two factors: L1 and L2 vocabulary, it is also important to note that general ability on some of the aforementioned variables (e.g., fluid intelligence, phonological storage) as possible confounds will be addressed in the subsequent regression analyses.

Measures

The study included group and individual administrations of a battery of tests. The series of tests were counterbalanced into one of six presentation orders. Reported below is a description of each measure. The sample reliabilities for each measure are reported in Appendix A. For commercial measures, reliabilities of the measures in the manual are also reported.

Math calculation. The arithmetic subtest from the Wide Range Achievement Test (WRAT-III; Wilkinson, 2003) was administered to measure basic calculation ability at first, second and third testing waves of the study. The WRAT-3 subtest required the child to perform written computation on number problems that increased in difficulty. The items vary from single-digit addition (2 + 2 = ?) to more advanced computational skills such as algebra. The WRAT-3 math calculation subtest allows up to 15 minutes for students to complete math calculations. The dependent measure was the number of problems correct (raw score range was 0-55), which yielded a standard score (M = 100, SD = 15). Internal consistency reported in the manual for the math subtest (Arithmetic - Blue Form) varied between .81 and .92 for the age ranges involved in the present study. The manual reported test-retest coefficients between .91 to .98.

Reading. The appropriate norm-referenced Woodcock reading tests were administered to the monolingual and ELL sample. The monolingual sample reading scores were assessed on the Woodcock-Reading Mastery test (e.g., Woodcock, 1998) whereas the bilingual sample was tested with the *Woodcock-* Muñoz *Language Survey-Revised (WMLS-R*; Woodcock, Muñoz-Sandoval, & Alverado, 2005). The subtests included word identification and passage comprehension. The dependent measure was the number of words identified or questions answered correctly. The manual reported test reliabilities that ranged from the mid-70s to high-90s; for the reading clusters.

Fluid intelligence. The Colored Progressive Matrices test (Raven, 1976) was used as an indicator of nonverbal problem solving or fluid intelligence. Children were given a booklet with patterns on each page, each pattern revealed a missing piece. For each pattern, six possible replacement pattern pieces were displayed. The dependent measure (range 0 to 36) was the number of problems solved correctly. The technical manual reports internal consistency reliabilities ranging from .80 to .90.

Vocabulary. Although the ELL sample was school identified on the CELDT measure, individual tests were administered by the research team that captured both expressive and receptive languages in both English and Spanish. The Expressive One-Word Picture Vocabulary Test–Spanish Bilingual Edition (EOWPVT-SBE) (Brownell, 2001) assessed an individual's English and Spanish speaking vocabulary. To administer the test the examiner presented a series of pictures of objects, actions or concepts to the examinee. The participant was asked to name each picture. Items are administered in both languages. The technical manual reported Cronbach alpha and split-half reliability coefficients that range from .92 to .97. To estimate receptive language proficiency for the ELL sample, the English Peabody Picture Vocabulary test (PPVT-4, Dunn & Dunn, 2007) and the Spanish Test de Vocabulario en Imagenes (TVIP) [Dunn, Lugo, Padilla & Dunn, 1986]) were administered. For the PPVT-4 task, children were presented with four pictures and asked to select the picture that matched the word read aloud in English. Item presentation gradually increased in difficulty. The technical manual reported a reliability of .91. To estimate Spanish vocabulary, the Test de Vocabulario en Imagenes (TVIP) was administered. This measure is similar to the PPVT in the presentation and administration, except that words are read aloud in Spanish. The split-half reliability presented in the technical manual varied between .91-.94.

As indicated from the school records, none of the monolingual children spoke Spanish nor was Spanish spoken in the home. However, to determine if English vocabulary proficiency within the monolingual sample was within the average range, we administered the vocabulary subtest from the WISC-III (Wechsler, 1991). Children were heard word by the examiner and were asked to provide the meaning of the word. The coefficient alpha for the current sample was .88.

Rapid naming of digits and letters. The administration followed those specified in the manual of the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 2000). For this task, the examiner presented the child with an array of items (e.g., letters or digits) on cards. Children were asked to name the items on the card, speaking in either English or Spanish, as quickly as possible for each stimulus set. The examiner used a stopwatch to time the children as they named all 72 items (2 sets of digits and 2 sets of letters). The dependent measure was the total combined time it took for students to complete each set. The manual reports correlations between parallel forms ranging from .80 to .93.

Inhibition. A random generation task was administered to assess inhibition**.** Each child was asked to write numbers (or letters) as quickly as possible, first in sequential order and then out of order. For example, children were first asked to write the numbers from 0 to 10 in order (i.e., 1, 2, 3, and 4) as quickly as possible in a 30-second period. They were then asked to write numbers as quickly as possible "out of order" within a 30-second period. Likewise, children were asked to write the letters of the alphabetic in order within 30-second period, and then were asked to write letters of the alphabetic out of order within a 30-second time period. Scoring included the number of numbers or letters that were random. Randomness was evaluated on three factors: seriation, repetition, and cycling. Seriation included random generation indices that tapped an inability to suppress stereotypical schemas such as repeated patterns (e.g., 1-4,….1-4; A-D,…A-D) and consecutive ascending and descending counting sequences (e.g., 4-5; C-D). The repetition factor included responses using same number or letter in succession (e.g., 3-3; B-B). Cycling was measured as the number of repetitions of the same number or letter within a sequence.

Working Memory Measures

Short-term memory (STM) measures (phonological loop). Four measures of STM were administered in English: Forward Digit Span, Backward Digit Span, Word Span, and Pseudoword Span. The Forward and Backward Digit Span task (taken from the WISC-III; Wechsler, 1991). The Forward Digit Span task required children to recall sequentially ordered sets of digits that increased in number, which were spoken by the examiner. The technical manual reported a test-retest reliability of .91. The Backward Digit Span task required children to recall sets of numbers, but in reverse order. [3] The reliability reported for this task was .76. Dependent measures for both tasks were the largest set of items recalled in order (range = 0 to 8 for Digits Forward; range = 0 to 7 for Digits Backward).

The Word Span and Pseudoword Span tasks were presented in the same manner as the Forward Digit Span task (Swanson & Beebe-Frankenberger , 2004). 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 (Nonword 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). As shown in Appendix A, the sample reliability for this task was .86.

Executive component of WM. A Conceptual Span, Listening Sentence Span, and Updating task were administered to capture the executive component of WM. Previous studies have shown that these measures load on the executive component of WM (seeSwanson, 2015; 2017). The WM tasks required 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 the Daneman and Carpenter's (1980) seminal WM measure.

The Conceptual Span task 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 a distracter question (which word did I say ‘shoes” or “socks”?) 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. The dependent measure for both versions was the number of items in the sets recalled correctly (range 0 to 32).

The children's adaptation of Daneman and Carpenter's (1980) Listening Sentence Span task was also administered (Swanson et al., 2008). 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 was also administered. A series of one-digit numbers was presented that varied in set length from 3, 5, 7, and 9. 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 long. 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 numbers, 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 (Swanson, 2013). The Visual Matrix task assessed the ability of participants to remember visual sequences within a matrix. Participants were presented with 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 the 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 of 0 to 11).

The Mapping and Directions task required the child to remember a sequence of directions on a map. The experimenter presented a street map with dots connected by lines; the 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 with 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 - 9) and in which the distracter process question was also answered correctly.

Procedures

Children were tested individually and in groups after informed consent forms were obtained for participation. For each testing wave, two sessions of individual testing were conducted, lasting thirty to sixty minutes for each session. Group testing occurred over the course of two consecutive days for approximately one hour each day. One of six presentation orders related to the individually administered tasks (WM, STM, phonological processing, and reading) was randomly assigned to each child.

Design

The design consisted of making three remeasurements of cross-sectional age groups (three cohorts) so that overlapping measurements of older and younger participants were provided. For the first testing wave, monolingual and ELL children in grades 1, 2 and 3 were tested on the aforementioned battery of academic, language and cognitive measures. For the second year of the project (wave 2), those same participants were retested on the same measures in the fall 1 year later. For year three (wave 3), participants who were in grades 2, 3, and 4 of the previous year were again retested in the fall in Grades 3, 4 and 5. Figure 1 shows the mean performance on the calculation measure for each cohort as a function of the monolingual and ELL participants. As shown, there is a general linear trend in computation growth across age. However, some cohorts show less change from grade to grade suggesting that nonlinear growth will need to be modeled in the complete regression model.

Approach to Analyses

The results are organized into three parts. First, we determined if our categorization of the tasks provided a good fit to the data. Thus, a confirmatory factor model was computed on the first testing wave. This was done for measurement purposes (latent variables control for measurement error as different variables have different weightings on a construct) and also for practical reasons: some constructs (e.g., WM, STM) included several tasks. We used the SAS CALIS (2010) program to create factor scores (latent variables) for each set of measures with two or more variables. This procedure allowed us to calculate standardized beta weights. Task weightings for the latent measures used in the analysis (as well as means and *SD*s and task sample reliabilities at wave 1) are presented in Appendix A. Measures met standard criteria for univariate normality with skewness for all measures less than 3 and kurtosis less than 4. Multivariate outliers were examined by calculating Mahanalobis' d^2. None of the cases were deemed outliers.

Based on the standardized loadings in Appendix A, latent scores were computed by multiplying the z-score of the target variable by the standardized factor loading weight based on the total sample at wave 1 (see Nunnally & Bernstein, 1994, p. 508 for calculation procedures). Consistent with previous investigations of growth (e.g., Wilson et al., 2002), all measures at wave 1 were scaled to have a mean of 0 and a standard deviation of 1. Wave 2 and 3 measures were z-scores based upon the means and standard deviations, as well as factor weightings, of wave 1. This was done across the total sample so that all parameters were on the same metric, thus enabling meaningful comparisons across testing waves (see McGraw & Joreskog, 1971, for a discussion).

Second, a mixed regression analyses determined whether growth in the aforementioned variables (i.e., reading, fluid intelligence, STM, naming speed, inhibition, WM) was related to growth in math computation. The formula (see Mehta & West 2000; Siliwinski, Hoffman & Hofer, 2010) on the criterion (math computation) measures can be shown as:

$$\text{Math computation}_{ti} = \beta_{0i} + \beta_{1i}(\text{time}_{ti}) + \beta_{2i}(\text{time}_{ti} \times \text{time}_{ti}) + e_{ti} \qquad (1)$$

$$e_{ti} \sim N(O, \sigma^2)$$

Where a child's math ability (*i*) at time *t*, math computation $_{ti}$, is a function of an individual-specific intercept parameter (β_{0i}), and (β_{1i}) is the individual-specific linear slope parameter, and (β_{2i}) is the, individual-specific nonlinear slope parameter, and the residual error (e_{ti}). For all models, the intercept (β_{0i}) reflected the average sample performance on the criterion measure (i.e., math) at wave 3. In contrast, the linear growth or slope parameter β_{1i} , reflected growth as a function of the chronological distance between testing occasions, which is specific to a given individual rather than to a given age. The mixed regression analyses included computing a baseline model (unconditional means model that included covariates) and four conditional models. The baseline model took into consideration individual linear and curvilinear growth effects.

For cohort-sequential studies, it is recommended to set the slope at the age in which the cohorts overlap (e.g., Mehta & West, 2000; see p. 29 for discussion**).** In this case, the three cohorts overlapped at grade 3 or age 9. Thus, β_{1i} and β_{2i} estimated growth (age-related change) in terms of per unit change centered at age 9 for a given child across each month of the testing waves.

The first conditional model entered latent measures assumed to be related to the executive system. The latent measure included the executive component of WM, visual-spatial WM and inhibition (random generation). A latent measure of visual-spatial WM was entered into the model because of its association with executive processing (e.g., Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). This model tested whether growth in executive processing was related to growth in computation. Also entered into this model were variables not controlled for in the sample selection. The variables serving as covariates were gender (female=1, male=-1), Anglo ethnicity (+1 Anglo vs. other =0), Hispanic ethnicity (+1 Hispanic vs. other=0), SES (participated Federal lunch program=-1 vs. no participation=1), ELL status (CELDT administered = -1 vs. not administered +1), proficient bilingual (proficient bilingual=+1, less proficient bilingual= -1, monolingual= 0), bilingual advantage (proficient Bilingual=+1, monolingual=-1, less proficient bilingual=0).

Also to examine whether wave 3 performance was related to age-related changes as a function of grade level , the interaction of cohort x age (predictor) variables for predicting the intercept (criterion measure at wave 3) was entered into the model at wave 1.

The age x cohort interaction was entered into the regression since it could be argued that the outcome variable is influenced by the instructional emphasis within each grade (i.e., early grades emphasize memorizing math facts whereas later grades emphasize word problem-solving).We assumed that the cohort effect was partially reflected by the grade level at which children were first tested at wave 1. However, it is important to note that to simplify to the analyses, we did not test for cross factors (influence of children moving for grade to grade).

The remaining conditional models tested whether the relationship between the executive component of WM and computation was significant when competing measures (e.g., reading, STM) were entered into the model. To examine the extent to which between-child variance in WM and math performance was influenced by latent measures of fluid intelligence and reading, our second conditional model entered measures assumed to capture the child's knowledge base (reading, fluid intelligence). The third model captured the influence of latent measures of English STM and naming speed in partialing out the influence of the executive component of WM on computation performance. The fourth model entered all the variables into a full regression model.

As shown in Table 3 and 4, the full mixed model represented the intercept at wave 3 and covariates at wave 1 and the linear and nonlinear slopes centered at grade 3 in the form:

$\beta_{0i} = \gamma_{00} + \gamma_{01}$ (Age at wave 1) + γ_{02} (age x cohort) + γ_{03} (gender) + γ_{04} (Hispanic ethnicity-1) + γ_{05} (Anglo-ethnicity-2) + γ_{06} (SES) + γ_{07} (CELDT vs. non CELDT administration) + γ_{08} (bilingually proficient vs. less proficient Ells) + γ_{09} (monolingual vs proficient ELLs) + γ_{010} (fluid intelligence) + γ_{011} (reading) + γ_{012} (STM) + γ_{013} (Naming speed) + + γ_{013} (Exec-WM) + γ_{013} (Visual-WM) + γ_{013} (Inhibition) + u_{0i},

$\beta_{1i} = \gamma_{10} + \gamma_{11}$ (reading) + γ_{12} (fluid intelligence) + γ_{13} (STM) + γ_{14} (naming speed) + γ_{15} (executive WM) + γ_{16} (visual-WM) + γ_{17} (inhibition) + $u_{1i.}$,

$\beta_{2i} = \gamma_{20} + \gamma_{11}$ (reading) + γ_{22} (fluid intelligence) + γ_{23} (STM) + γ_{24} (naming speed) + γ_{25} (Exec-WM) + γ_{26} (visual WM) + γ_{27} (inhibition) + $u_{2i.}$

All models were fit to the data using the SAS PROC Mixed software (SAS, 2010). Maximum likelihood (ML) procedures were used to determine the parameter estimates and account for missingness (see Peugh & Enders, 2004, for discussion). To account for the influence of children nested within classrooms, the multilevel regression model initially included as random effects for children's assignment to the various classroom/teachers at waves 1, 2, and 3. However, this complete cross-classification model (nested effects for wave 1, 2, and 3; see Hox, 2010) would not converge. Thus, for parsimony, only the random effects at wave 1 are reported since this testing wave had the largest intraclass correlation. A separate model testing occurred using the nesting at wave 2 and 3. However, the pattern of the results was comparable and therefore random effects included classroom/teacher at wave 1. Robust standard errors (Huber-White) were computed to allow for the nonindependence of observations from children nested within classrooms. To facilitate interpretation findings, all explanatory measures were grand mean centered.

The conditional models were compared to each other, as well as to an unconditional mean model and a demographic model that directly compared ELL and monolingual children. This comparison was done by determining the differences between the deviance values (i.e., the likelihood value for the correspondence between model and data), Akaike's Information Criterion (AIC) values, and Bayesian Criterion (BIC) values from the unconditional (baseline) and conditional growth model. In general, models with lower deviance, AIC and BIC values fit better than models with higher values.

Results

Normative Sample Representation

Table 1 shows the means and standard deviations on normed referenced measures of math computation, fluid intelligence, English word identification, English reading comprehension, and vocabulary scores for the monolingual, bilingually proficient and less bilingually proficient ELL children for each testing wave. Overall, the mean norm-referenced scores show the sample was within the average range across the three testing waves on measures of English vocabulary, English word identification, and fluid intelligence. Lower scores occurred on measures of math computation and English reading comprehension when compared to other normative measures. The median scores for the proficient bilingually proficient ELLs were in the normal standard score range, whereas median scores for less bilingually proficient ELLs were below the normal range (< 85 standard score).

Data Preparation

Confirmatory factor analysis. The factor structure and individual task loadings for measures used in the analysis are shown in Appendix A. Also reported in Appendix A are the sample reliability, means, and standard deviations for each task at wave 1. The comparative fit index (CFI), Bentler-Bonett Non-normed index (NNFI) and standardized root mean square error of approximation (SRMSEA) were computed to determine if the a priori categorization of measures fit the data. Values at .90 and over on the CFI and NNFI and RMSEA values of .05 or less indicate an acceptable fit. The model provided a good fit to the data (CFI = .99, RMSEA=.02 SRMSEA = .05, and NNFI = .99, $\chi^2(70)=77.77$, $p=.25$.).

Table 2 provides the means and SD for math and latent measures for each testing wave. Also provided are the effect sizes comparing the three groups (monolingual, bilingually proficient ELLs, less bilingually proficient ELLs) at each testing wave. Using Cohen's (1988) criteria for a medium and large effect sizes (absolute ES > .50 and .80, respectively), Table 2 shows the magnitude of differences between monolingual and ELL children on math calculation performance became increasingly apparent when testing children at wave 2 and 3. The large effect size differences between monolingual and ELL children occurred on measures of STM and the executive component of WM across all three testing waves. However, these findings must be qualified since preexisting difference (covariates discussed below) were not entered into the analysis. The contribution of preexisting differences among the ELL and monolingual children and the contribution of the executive system of WM to math computation growth are evaluated in the next analysis.

Growth Modeling

Unconditional means model. The growth models for the total sample are shown in Tables 4 and 5. Table 4 reports the fixed effects and Table 5 reports the growth effects and random effects for the unconditional means model. For Table 4, estimates of the intercept for computation at wave 3 are reported. For example in Table 4, the average z-score for computation at wave 3 for 9-year-olds in the total sample for the unconditional means model was .56. That is, the average intercept across children (the average value at wave 3 for children in grade 3 or age 9) was a z-score of .56. Because wave 3 scores were based on the mean and standard deviations of Wave 1, the average level of performance for 9-year-olds at Wave 3 showed an estimated improvement of approximately over 1/2 standard deviation from wave 1. In addition, the linear growth parameter reported in Table 4 ($\beta=.13$) (the average increment in computation per age in months) was significant. The nonlinear growth parameter was negative and significant, suggesting that there was a bend in the trajectory of computation growth. Hence, the important findings were that the average 9-year-old child at wave 3 yielded a z-score of .56 and gained .13 points per increase in age (per month).

The random effects portion of the model is shown at the bottom of Table 4. The random effects (error) for the unconditional means model included the intercept variance between classrooms (.20), the covariance between the intercept and slope (.006), the intercept variance between classrooms in terms of growth rate (.0003) and the residual error (.22). The intraclass correlation for the unconditional means model in predicting computation was .48 [.20 + .0003/.20 + .0003 + .22).

At the bottom of Table 4, the model fit indices are reported. The Likelihood value (deviance) indicated how well the model fits the data, the AIC allowed for a comparison of models that are not nested, and the BIC allowed for comparison of nested models (Hox, 2010, pp. 47-50). Lower Fit indices yield a better fit to the data than high fit indices.

Demographic model. In addition to the unconditional means model, we also computed a demographic model that took into consideration preexisting differences between ELL and monolingual children in predictions of math computation. As shown in Table 3, a number of significant differences emerged at wave 1 that contributed significant variance to Wave 3 computation performance. Significant parameters occurred for the cross-sectional age effect (older participants' out-performed younger participants) and the age x cohort interaction (older cohort performed better than the younger cohort). Also significant were variables related to SES and proficient bilingualism. Children participating in a Federal lunch program at wave 1 yielded significantly lower math computation scores three years later than children not participating in the lunch program. In addition, Wave 1 children who were relatively balanced in English and Spanish performance at wave 1, clearly out-performed ELLs who were less proficient bilinguals in computation at wave 3. The results also indicated there were no significant differences between monolingual and bilingually proficient ELLs as a function of being identified on the CELDT in later math computation performance.

Conditional models. Tables 3 and 4 also shows the results for four conditional models that included the demographic variables and executive processing (Model 1), reading and fluid intelligence measures (Model 2), phonological storage and naming speed measures (Model 3) and the full model (Model 4) in the growth modeling. These explanatory variables were entered simultaneously into the mixed regression analyses.

Model 1 yielded a number of significant effects when predicting Wave 3 math computation performance. As expected, the cross-sectional effect was significant (older participants out performed younger participants) as well as the age x cohort interaction (grade levels) as predictors of wave 3 math computation. The results also indicated that computation performance at wave 3 was significantly related to SES in that children **not** participating in the Federal lunch yielded higher math computation scores than those participating in the Federal lunch program. Wave 1 performance on all three executive processing measures predicted wave 3 computation performance. In terms of growth, the slope for the executive component of WM was significantly related to growth in computation. No other significant parameter estimates emerged.

When compared to the demographic model, the parameter estimate for Model 1 for the proficient vs. less proficient bilingual variable was no longer significant. By entering measures of executive processing into Model 1, potential advantages in bilingual proficiency in predicting later math performance was reduced by almost 50% ([.20 - .10]/.20). These findings support the notion that executive processing plays a major role in partialing out the role of bilingual proficiency in later math performance.

Model 2 tested whether growth effects related to math and the executive component of WM were partialed out by entering Fluid/Crystallized variables into the analysis. For the fixed effects, significant parameters emerged for Wave 1 measures of the cross-sectional effect of age, the age x cohort interaction, reading, fluid intelligence and the executive processing variables in uniquely predicting wave 3 math computation. As found in Model 1, a significant effect did not emerge for SES in the prediction of math computation at wave 3. A significant effect occurred for the bilingually proficient ELL children which indicated that proficient bilinguals (ELLs) yielded higher math scores at wave 3 than less bilingually proficient ELLs. In terms of growth, the only significant parameters related growth were measures of reading and the executive component of WM. The important finding was that measures of fluid intelligence and reading did not partial out the relationship between growth in WM and growth in computation.

Model 3 tested whether growth effects related to math and the executive component of WM were partialed out by entering measures of STM and naming speed into the analysis. For the fixed effects, significant parameters emerged for Wave 1 measures of age, age x cohort interaction, STM, naming speed and the executive processing variables uniquely predicted wave 3 math computation. Also predictive of Wave 3 computation were Wave 1 measures of SES. In terms of linear growth, the only significant parameters related to computation growth were measures in naming speed and the executive component of WM. In terms of nonlinear growth, the parameter estimate for computation growth was significantly related to growth in naming speed and the executive component of WM. The important findings for this model were that measures of STM and naming speed did not partial out the relationship between growth in WM and growth in computation.

As shown in Table 4 and 5, the Full model (Model 4) entered all the explanatory variables into the analysis. For the fixed effects, Wave 1 measures of age (cross-sectional effect), reading, fluid intelligence, naming speed and the executive processing variables in Wave 1 uniquely predicted wave 3 math computation. In contrast to the other conditional models, no significant age x cohort effect emerged. In terms of linear growth, math computation was related to linear growth in naming speed and the executive component of WM. In terms of nonlinear growth, growth in math computation was significantly related to growth in naming speed and growth in the executive component of WM. The deviance, AIC and BIC indices were lower for the Full Model, indicating a better fit to the data than Conditional Models 1,2, and 3.

In a follow-up analysis, we determined whether the full model captured a substantial amount of the variance in predictions of computation. In general, the full model accounted for a 60% ([.20- .08]/.20) reduction in the intercept variance when compared to the unconditional means model. Because previous studies have indicated that WM failed to provide significant variance in predictions of computation when competing measures (e.g., reading, fluid intelligence, STM) were entered into the regression analysis, the reduction in parameter estimates was computed. Thus, we compared the reduction in parameters related to WM in Model 1 with parameter estimates in Model 4. Compared to Model 1, the full model yielded a 55% (.09 -.04 /.09) reduction in the executive WM parameter at wave 1 when predicting wave 3 computation and a 33% (.03 - .02/.03) reduction in the linear slope. Thus, although the executive component of WM uniquely predicted math computation, some of its variance was related to measures of reading, fluid intelligence, STM and naming speed.

Summary

The important findings related to the mixed regression analyses were that growth in the executive component of WM and naming speed were significantly related to growth in computation. The results also indicated that monolingual and ELL children defined at wave 1 were statistically comparable in computation at wave 3. In addition, the results showed that the significant advantages in calculation performance of bilingually proficient ELL children when compared to less proficient ELL children found in the demographic model were eliminated when measures of phonological storage and the executive component of WM were entered into the regression models.

Discussion

This study investigated the cognitive processes that predict growth in math calculation in monolinguals and ELLs alike. Three important findings emerged. First, along with naming speed, the results suggest that growth in the executive component of WM component was significantly related to growth in calculation performance. Second, performance on measures of reading, fluid intelligence, naming speed and executive processes in wave 1 were significantly related to wave 3 math computation performance. Finally, although significant cross-section effects, linear, and nonlinear growth effects emerged across all conditional models (older children outperform younger children in computation), no significant variables related to monolingual or ELL children comparisons emerged in predictions of later math computation performance. Thus, when the results are taken together, we found weak support for the notion that the executive component of WM was unrelated to math computation or that ELL children experience a math achievement gap when compared to monolingual children when measures of executive processing are taken into consideration. Before discussing the results further, we briefly review the models that directed the study.

Three models were tested. Each model was assumed to partial out the influence of the executive component of WM on growth in math computation. The first model suggests that measures of fluid and crystallized intelligence partial out the influence of WM on computation performance. Measures of fluid and crystallized intelligence in this study were related to nonverbal problem solving (fluid intelligence) and reading. Because some of the functions of the central executive system are to access to the reservoir of information stored in LTM (e.g., Baddeley & Logie, 1999), and, therefore, this model suggests that controlling for the activation of information from LTM (e.g., reading) would partial out any influence of WM on calculation performance. The results show that measures of reading and fluid intelligence at wave 1 were significant predictors of later math performance in the regression analysis, but the entry of such variables into the analysis did not partial out the influence of WM components on calculation performance. In addition, none of the measures of fluid and crystallized intelligence growth in the Full model were related to growth in math calculation.

A second model tested whether growth related to the phonological storage (phonological loop) eliminated the influence of growth on the executive component of WM on growth in math computation. Latent measures of STM and rapid naming speed captured phonological processing in this study. Weak phonological processing is viewed as a bottleneck to various aspects of children's executive processing and math performance (e.g., Hecht et al., 2001; Peng et al., 2016). The results show that measures of STM and naming speed were significant predictors of calculation in the later grades in the regression analysis for Model 3, but the entry of such variables into the analysis did not partial the influence of the executive component of WM on calculation performance. In addition, only growth in naming speed (not STM) was found to be significantly related to growth in computation.

The third model assumes that growth in executive processing was related to math computation. However, we assumed that the influence of executive processing within the WM system was related to inhibition. As shown in Model 1, inhibition was a significant predictor of later math performance, but its simultaneous entry into the regression model did not eliminate the contribution of WM in predicting later computation. It is also important to note that included in this model were measures of visual-spatial WM. Visual-spatial WM was included in the analysis because of its association with arithmetic difficulties (e.g., Ashkenazi et al., 2013; Crocker, Riley, & Mattson, 2015) and the executive component of WM (e.g., Miyake et al., 2001) and random generation measures (latent measures of inhibition) were included because of their association with executive processing (e.g., Cooper, 2016; Towse & Cheshire, 2007). The results in Model 1 showed that simultaneous entry of these measures measured at wave 1 were each significant predictors of math computation in the later grades. However, the inclusion of inhibition and visual-WM in the model did not partial out the significant influence of the executive component of WM in later math performance. The full regression model showed that growth in the executive component of WM was found to be significantly related to growth in computation even when competing measures of executive processing (inhibition, visual WM), STM and fluid and crystallized intelligence were entered into the model. The growth modeling yielded a positive parameter for the executive component of WM suggesting that increases in the executive component of WM were related to increases in math calculation. We now consider our first question:

Question 1: Is growth in math computation related to growth in the executive component of working memory, independent of

measures of fluid/crystallized intelligence, STM, naming speed and inhibition?

We find support for the notion that growth in the executive component of WM contributed unique variance to growth in math computation. However, we would not completely discount the role of fluid/ crystallized intelligence, STM or naming speed in accounting for the relationship between WM and computation growth parameters. Clearly, each model reduced the influence of the executive component of WM parameter by 33% to 55% in predictions of computation. We did not find support for the notion, however, that measures of fluid and crystallized measures of intelligence, inhibition or STM completely eliminated the influence of the executive component of WM on computation growth. Instead, we found that the influence of the executive component of WM variable on computation growth operated independent of fluid/crystallized intelligence, inhibition as well as STM measures.

Question 2: Are there math computation advantages for ELL children who are relatively proficient bilinguals than less proficient ELL or monolingual children?

This question focused on the math achievement gap commonly found in the literature comparing monolingual and ELL children. We did not find support in the Full model for the notion that ELL elementary school children are deficient in later math performance when compared to monolingual children. None of our dummy or contrast variables (e.g., SES, school identification as ELL, ethnicity, proficient bilingualism) comparing monolingual and ELL children were significant in the Full model. In addition, the bilingual proficiency (proficient vs. less proficient ELL children) advantages were eliminated when measures of executive processing were entered into the regression model. Thus, the results support the notion that there are performance advantages for relatively bilingually proficient ELL children and these advantages (math computation) are related to executive processing. In addition, these advantages occur for ELL children even though they show greater economic risk than the monolingual sample. Several studies have suggested a strong link between bilingualism and cognitive control over and above the effects of SES (Krizman, Skoe, & Kraus, 2016; Nair, Biedermann & Nickels, 2017). Our findings are consistent with studies that have shown that proficiency in L1 and L2 positively influences intentional control, cognitive flexibility and overall executive processing that in turn influences academic achievement (e.g., Han, 2012; Oades-Sese, Esquivel, Kaliski, & Maniatis, 2011). Thus, the present study suggests that bilingualism may have provided a protective factor in ELL children's academic achievement.

Competing Models

No doubt, there are several competing interpretations as to whether the executive component of WM underlies later computation performance; three are considered. First, naming speed mediates the role of WM growth on math computation. Several models assume that operations related to academic achievement (e.g. math) are time-related (e.g., Georgiou, Tziraki, Manolitsis, & Fella, 2013). Although performance on the naming speed tasks played a significant role related to growth effects for math computation in our study, we did not find support for the hypotheses that the relationship between growth in executive processing and growth in computation was mitigated when measures of speed were entered into the analysis.

Second, individual differences in executive processing are merely an artifact of individual differences referred to in Baddeley's model as the phonological loop (STM). That is, one of the components of WM commonly attributed to computation performance is the phonological loop. This model suggests that the phonological system has a bottom-up influence on WM. Because the phonological loop (STM) has been found to play a major role in math and math performance is highly correlated with math computation in the younger grades, it has been argued that the influence of higher order processes, such as executive processing, in predictions of computation is an artifact of phonological storage (e.g., Peng et al., 2016). A key finding of this study, however, was that growth in computation was significantly related to growth in WM performance even when specific processes (STM, inhibition) or more general processes (naming speed) were entered into the analysis.

A third possibility is that resistance to interference between the two language systems underlies the influence of WM on math performance. This model suggests that interference may be related to the competing influence of Spanish on math measures and/or a general ability to suppress competing linguistic information. As a consequence, ELL children may have trouble preventing unnecessary information in the Spanish language system from entering memory on English measures. This interpretation does not appear to be a viable alternative to the present results, for a number of reasons, the most obvious reason being that no significant differences in math computation as a function of ELL and monolingual children in the full conditional model emerged. There are three other reasons to consider (also see Swanson et al., 2015).

First, growth effects for the latent measures of inhibition growth were not significant covariates of growth on the computation measure. Thus, our measure of inhibition did not play a major role in explaining individual differences in math computation. Second, simultaneous entry of inhibition (as measured by the random generation task) into the mixed regression analysis did not partial out the influence of WM on predictions of computation.

A final reservation, and perhaps more importantly, the random generation measure may reflect different executive processing activities than the executive component of WM. The executive component of WM may be more closely aligned with updating than inhibition activities reflective of the random generation task. Thus, updating may be more closely aligned with computation growth in both the ELL and monolingual than inhibition.

An important question emerges as to what develops in ELL children's math computation? Two mechanisms were found to be important in the development of computation. The results clearly showed that growth in naming speed and the executive component of WM were the only measures that were significantly and uniquely related to computation growth in the current sample. Although these measures share an association (*r*s < .30 across each testing wave in the current sample), growth in each of these latent measures appears to have a unique function in computation development. Although we can only speculate on the mechanisms that underlie each of these measures, some inferences can be drawn from the literature. For example, naming speed influences phonological STM storage. Naming speed has been considered a measure of how quickly items can be encoded and rehearsed (e.g., Bonifacci et al., 2011; Georgio et al., 2013; McDougall, Hulme, Ellis, & Monk. 1994). That is, to reduce the decay of memory items in the phonological store prior to output, naming speed influences the effectiveness of subvocal rehearsal processes (e.g. Henry & Millar, 1993).

In contrast, the development of the executive component WM is related to the preservation of information while simultaneously processing other information and attention control related to these functions (e.g., Baddeley, 2012; Unsworth & Engle, 2007). Although the development of the executive components of WM tasks may share the same processes as STM (e.g., rehearsal, updating, controlled search), these tasks have a greater reliance on phonological processes than WM tasks (see Unsworth & Engle, 2007, pp. 1045-1046, for a review). Thus, we infer that because WM tasks are highly associated with the control of attention as well as the scope and focus of attention (e.g., Unsworth & Engle, 2007); these processes play an important role in the development of computation in both ELL and monolingual children.

Limitations

There are at least three limitations to this study. First, the interaction of classroom instructional math activities on student outcomes (e.g., end-of-year performance in math) was not explored. We expect these relationships on math outcomes will vary according to the language of instruction (Spanish vs. English reading), specific instructional activities, and the child's initial status (e.g., low working memory). It is important to note in this study that WM and potential mediating variables were administered in the language of instruction: English.

Second, although our primary focus is on the relationship between WM and its effects on math performance, no doubt other variables (covariates) that were assumed to underlie growth in phonological STM and executive processing in WM should have been considered in the analysis. For example, we have not adequately assessed the role of visual–spatial processes in this study; therefore visual–spatial WM may be relevant to other types of mathematical performance (Li & Geary, 2017; Mammarella & Cornoldi, 2014). In addition, we have not investigated various components of number sense, such as mental representations for number comparison and approximate calculation.

Finally, the sample reflected sequential bilingualism (L2 follows L1 development) and therefore may not reflect bilingualism when two languages are learned simultaneously. It is important to note that the majority of these studies on executive processing and bilingualism have focused on children who learned L1 and L2 simultaneously. However, ELL children in U.S. public schools frequently represent children who learn L1 first and L2 later (as they enter school). Thus, few studies have focused on sequential bilinguals (who learn their L1 first, then L2 later) with different levels of language proficiency on executive processing and math.

Summary

Taken together, we interpret our findings as suggesting that growth in math computation is directly tied to growth in math computation and the WM system. The results suggest that when the effects of reading, fluid intelligence, naming speed and inhibition were partialed out, growth in the executive component of WM was related to growth in computation ability. The results also suggested preexisting differences between ELL and monolingual children in predictions of later math performance were mitigated in the full regression model.

References

Ashkenazi, S., Rosenberg-Lee, M., Metcalfe, A. W. S., Swigart, A. G., & Menon, V. (2013). Visuo–spatial working memory is an important source of domain-general vulnerability in the development of arithmetic cognition. *Neuropsychologia, 51*(11), 2305-2317. doi:/10.1016/j.neuropsychologia.2013.06.031

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

Baddeley, A. (2007). *Working memory, thought, and action*. New York, NY, US: Oxford University Press. doi:10.1093/acprof:oso/9780198528012.001.0001

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

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language-learning device. *Psychological Review, 105,* 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.

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

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), 93-100. doi: 10.1016/j.bandl.2008.09.001

Blair, C., Ursache, A., Greenberg, M., & Vernon-Feagans, L. (2015). Multiple aspects of self-regulation uniquely predict mathematics but not letter–word knowledge in the early elementary grades. *Developmental Psychology, 51*(4), 459-472. Doi:10.1037/a0038813

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish–Dutch children. *Journal of Experimental Child Psychology, 128*, 105-119. doi:10.1016/j.jecp.2014.06.007

Bonifacci, P., Giombini, L., Bellocchi, S., & Contento, S. (2011). Speed of processing, anticipation, inhibition and working memory in bilinguals. *Developmental Science, 14*(2), 256-269.

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

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*(3), 273-293. doi:10.1207/S15326942DN1903_3

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies* Cambridge University Press, New York, NY. doi: http://dx.doi.org/10.1017/CBO9780511571312 Retrieved from https://search.proquest.com/docview/618390556?accountid=14521

Cai, D., Georgiou, G. K., Wen, M., & Das, J. P. (2016). The role of planning in different mathematical skills. *Journal of Cognitive Psychology, 28*(2), 234-241. doi:10.1080/20445911.2015.1103742

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale,NJ: L. Erlbaum Associates.

Colom, R., Shih, P. C., Flores-Mendoza, C., & Quiroga, M. Á. (2006). The real relationship between short-term memory and working memory. *Memory, 14*(7), 804-813. doi:10.1080/09658210600680020

Colom, R., Flores-Mendoza, C., Quiroga, M. Á., & Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality and Individual Differences, 39*(5), 1005-1014. doi:10.1016/j.paid.2005.03.020

Cooper, R. P. (2016). Executive functions and the generation of "random" sequential responses: A computational account. *Journal of Mathematical Psychology, 73*, 153-168. doi:10.1016/j.jmp.2016.06.002

Cowan, R., Donlan, C., Shepherd, D., Cole-Fletcher, R., Saxton, M., & Hurry, J. (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *Journal of Educational Psychology, 103*(4), 786-803. doi:/10.1037/a0024556

Crocker, N., Riley, E. P., & Mattson, S. N. (2015). Visual-spatial abilities relate to mathematics achievement in children with heavy prenatal alcohol exposure. *Neuropsychology, 29*(1), 108-116. doi:/10.1037/neu0000094

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

David, C. V. (2012). Working memory deficits in math learning difficulties: A meta-analysis. *International Journal of Developmental Disabilities, 58*(2), 67-84. doi:/10.1179/2047387711Y.0000000007

de Abreu, Pascale M. J. Engel. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory, 19*(5), 529-537.doi:10.1080/09658211.2011.590504

Dunn, L. M., & Dunn, L. M. (2007). *The Peabody Picture Vocabulary Test-4.* NY: Pearson.

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

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science,11*(1), 19-23. doi: 10.1111/1467-8721.00160

Farnia, F., & Geva, E. (2011). Cognitive correlates of vocabulary growth in English language learners. *Applied Psycholinguistics, 32,* 711-738. doi:10.1017/S0142716411000038

Friedman, N. P., Haber, B. C., Willcutt, E. G., Miyake, A. Young, S. Corely, R. P. & Hewitt, J. K. (2007). Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychological Science, 18,* 893-900. doi:10.1111/j.1467-9280.2007.01997

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(1)*, 29-43.doi:10.1037/0022-0663.98.1.29

Fuchs, L. S., Gilbert, J. K., Powell, S. R., Cirino, P. T., Fuchs, D., Hamlett, C. L., Tolar, T. D. (2016). The role of cognitive processes, foundational math skill, and calculation accuracy and fluency in word-problem solving versus prealgebric knowledge. *Developmental Psychology, 52*(12), 2085-2098. doi:/10.1037/dev0000227

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

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*(4), 1343-1359. doi:10.1111/j.1467-8624.2007.01069.x

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology, 104*(1), 206-223. doi:/10.1037/a0025398

Geary, D. C., Nicholas, A., Li, Y., & Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: An eight-year longitudinal study. *Journal of Educational Psychology, 109*(5), 680-693. doi:/10.1037/edu0000159

Georgiou, G. K., Tziraki, N., Manolitsis, G., & Fella, A. (2013). Is rapid automatized naming related to reading and mathematics for the same reason(s)? A follow-up study from kindergarten to grade 1. *Journal of Experimental Child Psychology, 115*(3), 481-496. doi: 10.1016/j.jecp.2013.01.004

Gorman, B. K. (2012). Relationships between vocabulary size, working memory, and phonological awareness in Spanish-speaking English language learners. *American Journal of Speech-Language Pathology, 21*(2), 109-123. doi:/10.1044/1058-0360(2011/10-0063

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

Grégoire, J., & Desoete, A. (2009). Mathematical disabilieties—An underestimated topic? *Journal of Psychoeducational Assessment, 27*(3), 171-174. doi:http://dx.doi.org/10.1177/0734282908330577

Han, W. (2012). Bilingualism and academic achievement. *Child Development, 83*(1), 300-321. doi: 10.1111/j.1467-8624.2011.01686.x

Harvey, H. A., & Miller, G. E. (2017). Executive function skills, early mathematics, and vocabulary in head start preschool children. *Early Education and Development, 28*(3), 290-307. doi:/10.1080/10409289.2016.1218728

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computational skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology, 79*(2), 192-227. doi:10.1006/jecp.2000.2586

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.

Henry, L. A. & Millar, S. (1993). Why does memory span improve with age? A review of the evidence for two current hypotheses. *European Journal of Cognitive Psychology, 5,* 241-287.

Horn, J.L. & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. D. P.Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp. 53-91), New York: Guilford Press.

Hox, J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd Ed.) New York, NY: Routledge/Taylor & Francis.

Jahanshahi, M., Saleem, T., Ho, A. K., Dirnberger, G., & Fuller, R. (2006). Random number generation as an index of controlled processing. *Neuropsychology, 20*(4), 391-399. doi: 10.1037/0894-4105.20.4.391

Kempert, S., Saalbach, H., & Hardy, I. (2011). Cognitive benefits and costs of bilingualism in elementary school students: The case of mathematical word problems. *Journal of Educational Psychology, 103*(3), 547-561. doi:/10.1037/a0023619

Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. M. (2014). Involvement of working memory in longitudinal development of number–magnitude skills. *Infant and Child Development, 23*(1), 36-50. doi:/10.1002/icd.1834

Klesczewski, J., Brandenburg, J., Fischbach, A., Grube, D., Hasselhorn, M., & Büttner, G. (2015). Working memory functioning in children with poor mathematical skills: Relationships to IQ–achievement discrepancy and additional reading and spelling difficulties. *Zeitschrift Für Psychologie/Journal of Psychology, 223*(2), 83-92. doi:/10.1027/2151-2604/a000206

Krizman, J., Skoe, E., & Kraus, N. (2016). Bilingual enhancements have no socioeconomic boundaries. *Developmental Science, 19*(6), 881-891. doi:http://dx.doi.org/10.1111/desc.12347

Kudo, M., & Swanson, H. L. (2014). Are there advantages for additive bilinguals in working memory tasks? *Learning and Individual Differences, 35*, 96-102. doi: 10.1016/j.lindif.2014.07.019

Li, Y., & Geary, D. C. (2017). Children's visuospatial memory predicts mathematics achievement through early adolescence. *PLoS One, 12*(2) doi:/10.1371/journal.pone.0172046

Llosa, L. (2007). Validating a standards-based classroom assessment of English proficiency: A multitrait-multimethod approach. *Language Testing, 24*(4), 489-515. doi:10.1177/

Lonigan, C. J., Allan, D. M., Goodrich, J. M., Farrington, A. L., & Phillips, B. M. (2017). Inhibitory control of Spanish-speaking language-minority preschool children: Measurement and association with language, literacy, and math skills. *Journal of Learning Disabilities, 50*(4), 373-385. doi:/10.1177/0022219415618498

Macizo, P., Herra, A., Roman, P., & Marin, M.C. (2011). Proficiency in a second language influences the processing of number words. *Journal of Cognitive Psychology, 23,* 915-921. Doi:10.1080/20445911.2011.586626

Mammarella, I. C., & Cornoldi, C. (2014). An analysis of the criteria used to diagnose children with nonverbal learning disability (NLD). *Child Neuropsychology, 20*(3), 255-280. doi:/10.1080/09297049.2013.796920

Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment, 14*(3-4), 160-179. doi:10.1080/10627190903422906

McDougall, S., Hulme, C., Ellis, A., & Monk, A. (1994). Learning to read: The role of short-term memory and phonological skills. *Journal of Experimental Child Psychology, 58*(1), 112-133. doi:10.1006/jecp.1994.1028

McGraw, B., & Joreskog, K. G. (1971). Factorial invariance of ability measures in groups differing in intelligence and social economic status. *British Journal of Mathematical and Statistical Psychology, 24*, 154-168.

Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5(1),* 23-43. doi:10.1037/1082-989X.5.1.23

Menon, V. (2016). Working memory in children's math learning and its disruption in dyscalculia. *Current Opinion in Behavioral Sciences, 10*, 125-132. doi:/10.1016/j.cobeha.2016.05.014

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

Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology, 114*(2), 187-202. doi:http://dx.doi.org/10.1016/j.jecp.2012.09.002

Nair, V. K. K., Biedermann, B., & Nickels, L. (2017). Effect of socio-economic status on Cognitive control in non-literate bilingual speakers. *Bilingualism: Language and Cognition, 20*(5), 999-1009. doi:http://dx.doi.org/10.1017/S1366728916000778

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

National Assessment of Educational Progress (2004). Washington DC: US Department of Education

National Assessment of Educational Progress (2011). Achievement gap: How Hispanics and white students in public schools perform in mathematics and reading on the national assessment of educational progress. Washington DC: US Department of Education.

National Assessment of Educational Progress (2017). The condition of education (update 2017) Washington DC: US Department of Education.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd Ed.). NY: McGraw-Hill.

Oades-Sese, G., Esquivel, G. B., Kaliski, P. K., & Maniatis, L. (2011). A longitudinal study of the social and academic competence of economically disadvantaged bilingual preschool children. *Developmental Psychology, 47*(3), 747-764. doi: 10.1037/a0021380

Peña, E. D., Bedore, L. M., & Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders, 51*(2), 192-202. doi:10.1111/1460-6984.12199

Peng, P., Namkung, J. M., Fuchs, D., Fuchs, L. S., Patton, S., Yen, L., Hamlett, C. (2016). A Longitudinal study on predictors of early calculation development among young children at risk for learning difficulties. *Journal of Experimental Child Psychology, 152*, 221-241. doi:/10.1016/j.jecp.2016.07.017

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology, 108*(4), 455-473. doi:/10.1037/edu0000079

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology, 108*(4), 455-473. doi:/10.1037/edu0000079

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74, 525-556. doi:10.3102/00346543074004525

Polk, R. L. (2016). *Examining students with disabilities in a linked learning pathway* (Order No. AAI3725395). Available from PsycINFO. (1821806720; 2016-31152-133). Retrieved from https://search.proquest.com/docview/1821806720?accountid=14521

Powell, S. R., Kearns, D. M., & Driver, M. K. (2016). Exploring the connection between arithmetic and prealgebraic reasoning at first and second grade. *Journal of Educational Psychology, 108*(7), 943-959. doi:/10.1037/edu0000112

Ratiu, I., & Azuma, T. (2015). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology, 27*(1), 1- 11. doi:10.1080/20445911.2014.976226

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

Roberts, G., & Bryant, D. (2011). Early mathematics achievement trajectories: English-language learner and native English-speaker estimates, using the early childhood longitudinal survey. *Developmental Psychology, 47*(4), 916-930. doi:/10.1037/a0023865

Rosen, V. M., & Engle, R. W. (1997). Forward and backward serial recall. *Intelligence, 25*(1), 37-47.

Rosselli, M., Ardila, A., Lalwani, L. N., & Vélez-Uribe, I. (2016). The effect of language proficiency on executive functions in balanced and unbalanced Spanish–English bilinguals. *Bilingualism: Language and Cognition, 19*(3), 489-503.doi:10.1017/S1366728915000309

SAS Institute. (2010). *SAS/STAT software: Changes and Enhancements through release 9.3*. Cary, NC: SAS Institute Inc.

Sliwinski, M., Hoffman, L., & Hofer, S. M. (2010). Evaluating convergence of within -person change and between-person age differences in age-heterogeneous longitudinal studies. *Research in Human Development, 7(1),* 45-60. doi:10.1080/15427600903578169

Snijders, T., & Bosker, R. (1999). *Multilevel modeling: An introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage Press.

Swanson, H.L. (2013). Abbreviated Test of Working Memory. American Psychological Association, Washington DC: PsychTESTS.

Swanson, H. L. (2015). Growth in working memory and inhibition predicts literacy in English language learners: A cross-sectional and longitudinal study. *Memory, 23*(5), 748-773. doi:10.1080/09658211.2014.927504

Swanson, H. L. (2017). Verbal and visual-spatial working memory: What develops over a life span? *Developmental Psychology, 53*(5), 971-995. doi:10.1037/dev0000291

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., & Fung, W. (2016). Working memory components and problem-solving accuracy: Are there multiple pathways? *Journal of Educational Psychology, 108*(8), 1153-1177. doi:10.1037/edu0000116

Swanson, H. L., Jerman, O., & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 100*(2), 343-379. doi:10.1037/0022-0663.100.2.343

Swanson, H. L., Kong, J., & Petcu, S. (2018a). Math difficulties and working memory growth in English language learner children: Does bilingual proficiency play a significant role? *Language, Speech, and Hearing Services in Schools, 49*(3), 379-394. doi:10.1044/2018_LSHSS-17-0098

Swanson, H. L., Kudo, M. F., & Van Horn, M. L. (2018b). Does the structure of working memory in el children vary across age and two language systems? *Memory,* doi:10.1080/09658211.2018.1496264

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*(2), 321-333. doi: 10.1037/0022-0663.91.2.321

Swanson, H. L., Orosco, M. J., & Lussier, C. M. (2012). 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. J., & Lussier, C. M. (2015). Growth in literacy, cognition, and working memory in English language learners. *Journal of Experimental Child Psychology, 132*, 155-188. doi:10.1016/j.jecp.2015.01.001

Toll, S. W. M., Van, d. V., Kroesbergen, E. H., & Van Luit, Johannes E. H. (2011). Executive functions as predictors of math learning disabilities. *Journal of Learning Disabilities, 44*(6), 521-532. doi:/10.1177/0022219410387302

Towse, J., & Cheshire, A. (2007). Random generation and working memory. *European Journal of Cognitive Psychology, 19 (3),* 374-394. doi:10.1080/09541440600764570

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

Vukovic, R. K., & Lesaux, N. K. (2013). The language of mathematics: Investigating the ways Language counts for children's mathematical development. *Journal of Experimental Child Psychology, 115*(2), 227-244. doi: 10.1016/j.jecp.2013.02.002

Wagner, R., Torgesen, J., & Rashotte, C. (2000). *Comprehensive Test of Phonological Processes.* Austin TX: Pro-ED.

Weber, R. C., Johnson, A., Riccio, C. A., & Liew, J. (2016). Balanced bilingualism and executive functioning in children. *Bilingualism: Language and Cognition, 19*(2), 425-431. doi:10.1017/S1366728915000553

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

Wilkinson, G. S. (2003). *The Wide Range Achievement Test*. Wilmington DE: Wide Range, Inc.

Wilson, R. S., Beckett, L. A., Barnes, L., Schneider, J., Bach, J., Evans, D., Bennett, D. A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging, 17,* 179-193.

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

Footnote

1. The literature is unclear as to what terms appropriately capture our sample (e.g., English language learners, English learners, limited English-proficient, balance vs. unbalanced bilingual, emerging bilinguals). We used the term English language learner to align with the literature, but realize the sample is best described as bilingual or emerging bilingual. This is because the children varied in proficiency in their primary language and/or were weak in both language systems.
2. The CELDT assesses English learners in the following areas: listening, speaking, reading and writing. The test is given to students whose Home Language Survey indicates a language other than English is spoken in the home, and for whom there is no prior record of English language testing. The test is administered annually to track English proficiency. The test score yields a proficiency level represented numerically as 1-5. The five levels of the CELDT are: Beginning, Early Intermediate, Intermediate, Early Advanced, and Advanced. This measure has been validated as a measure of English Proficiency (e.g., Llosa, 2007, CDE, 2008) and because this was a public school classification of ELL students in our study, we assumed the measure had face validity.
3. Colom et al. (2005), Rosen and Engle (1997), Swanson, Mink and Bocian (1999) find that the numbers reversed task is a short-term processing capacity measure. For example, Rosen and Engle (1997) stated that "We found no evidence that backward recall required more complex recall than did forward recall. Also, we found that participants used a phonological code for both forward and backward serial recall" (p. 46). Other studies have found forward and backward recall of digits load on the same factor (e.g., Colom et al., Swanson et al., 1999). We did however, run the analysis without backward digit span and found that the pattern of results was comparable. Thus, in the present study, the backward span task as a measure of phonological STM was maintained and provided an excellent fit to the data.

Table 1

Norm Referenced Scores for Math, Fluid Intelligence, Reading, and Vocabulary for Monolingual and English Language Learners

					Wave 1				
		Monolingual (English)			ELL (Proficient bilingual)[a]			ELL (Less Proficient bilingual)[a]	
Variable	N	M	SD	N	M	SD	N	M	SD
AGE (mos.)	358	92.66	11.83	236	90.77	11.5	247	92.49	11.04
Calculation	358	111.38	12.78	236	105.02	11.62	247	100.64	12.96
Fluid Intell.	358	107.44	14.6	225	104.62	15.3	232	100.28	14.24
Word ID	351	106.44	15.05	234	106.07	14.07	247	97.53	14.83
Comprehension	351	104.82	11.55	236	95.18	12.85	247	86.34	15.4
Vocabulary-WISCIII	351	102.88	24.81						
PPVT				234	88.39	7.88	247	78.39	8.86
TVIP				234	90.49	13.14	246	73.51	14.83
E-Expressive			.	236	99.62	11.88	247	83.16	18.89
S-Expressive				236	74.67	16.79	247	61.83	13.32
Median				236	89.40	5.99	247	73.52	7.08
					Wave 2				
AGE (mos.)	321	105.37	11.61	213	102.77	11.5	213	104.49	11.04
Calculation	321	110.74	13.62	218	102.00	12.03	213	97.23	11.49
Fluid Intell.	321	109.42	12.83	213	106.61	11.53	212	101.09	12.88
Word ID	321	103.82	14.57	218	105.95	14.25	216	96.54	16.11
Comprehension	321	103.22	11.5	217	98.42	11.5	214	88.88	14.57
Vocabulary-WISCIII	321	109.56	26.93						
PPVT				217	88.87	10.17	215	81.33	11.02
TVIP				218	85.23	14.86	216	72.77	13.93
E-Expressive				218	103.42	12.17	216	91.91	16.51
S-Expressive				218	77.07	16.48	216	65.23	13.38
Median				236	88.82	9.67	216	76.23	9.86
					Wave 3				
AGE (mos.)	303	117.13	12.1	186	114.77	11.5	185	116.49	11.04
Calculation	303	111.36	13.67	186	101.35	12.43	185	95.71	13.3
Fluid Intell.	301	109.76	12.79	187	107.42	11.77	185	102.67	12.87
Word ID	303	104.78	15.94	188	103.73	16.02	183	93.48	16.5
Comprehension	303	101.22	11.02	188	94.08	13.11	184	86.73	13.9
Vocabulary-WISCIII	302	108.55	26.64						
PPVT				188	90.80	9.91	184	82.94	11.22
TVIP				188	84.48	17.52	184	69.69	13.69
E-Expressive				187	105.76	12.93	182	94.27	13.81
S-Expressive				187	77.37	15.66	182	65.30	12.03
Median				187	90.24	11.54	184	76.36	12.52

Note. [a] Bilingual Proficiency was defined as median of total score for English and Spanish receptive and expressive vocabulary above an 85 norm- referenced standard score whereas children designated with less bilingually proficient yield median vocabulary scores at or below an 85 norm-referenced standard score. Word ID=word identification, Fluid Intell=Fluid intelligence-Raven Colored Matrices Test, Comprehension: Passage Comprehension TVIP =Test de Vocabulario en Imagenes, E-expressive =English section of EOWPVT-SBE, S-Expressive=Spanish section of EOWPVT-SBE, Median=median of standard score for PPVT-4, TVPT, and EOWPVT-SBE.

Table 2

Z-score Comparisons on Math, Reading, Fluid Intelligence and Cognitive Measure

	Monolingual			ELL-Proficient			ELL-less proficient			Effect Sizes	
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	1 vs. 2	1 vs. 3	2 vs. 3
Chronlogical Age												
Age (Mos)	358	92.53	11.83	237	90.74	11.46	250	92.47	11.03	0.15	0.01	-0.15
Math Computation												
Wave 1	358	0.12	1.08	236	-0.03	0.94	247	-0.15	0.92	0.15	0.27	0.13
Wave 2	322	1.08	1.19	214	0.37	0.9	214	0.23	0.9	0.68	0.81	0.16
Wave 3	303	1.83	1.02	186	0.91	0.86	186	0.78	0.97	0.98	1.06	0.14
Reading												
Wave 1	351	0.65	1.26	237	0.01	1.29	250	-0.92	1.45	0.50	1.16	0.68
Wave 2	323	0.44	1.26	217	0.21	1.23	215	-0.77	1.50	0.18	0.88	0.72
Wave 3	303	0.35	1.23	187	-0.16	1.38	184	-1.05	1.47	0.39	1.04	0.62
Fluid Intelligence												
Wave 1	358	0.11	0.97	226	0.03	1.03	232	-0.2	0.98	0.08	0.32	0.23
Wave 2	322	0.65	0.86	213	0.59	0.83	214	0.31	0.93	0.07	0.38	0.32
Wave 3	303	1.01	0.75	187	0.98	0.75	187	0.77	0.78	0.04	0.31	0.27
Phonological Storage (STM)												
Wave 1	353	1.58	1.22	235	-0.92	1.35	248	-1.36	1.41	1.95	2.24	0.32
Wave 2	323	2.38	1.21	217	0.26	1.27	215	-0.21	1.26	1.71	2.1	0.37
Wave 3	302	2.05	1.22	188	0.81	1.28	185	0.54	1.34	0.99	1.18	0.21
Naming Speed												
Wave 1	353	-0.11	1.32	237	0.07	1.47	247	0.08	1.34	-0.13	-0.14	-0.01
Wave 2	322	-0.76	0.86	218	-0.38	1.15	214	-0.29	1.29	-0.38	-0.44	-0.07
Wave 3	303	-1.11	0.7	188	-0.65	1.04	184	-0.56	1.03	-0.53	-0.64	-0.09
Exec-WM												
Wave 1	354	1.08	1.49	239	-0.5	1.24	250	-0.88	1.03	1.16	1.56	0.33
Wave 2	323	2.23	1.4	218	0.38	1.08	216	-0.13	0.92	1.49	2.03	0.51
Wave 3	303	2.59	1.21	183	0.68	1.36	178	0.24	1.07	1.49	2.06	0.36
Vis-WM												
Wave 1	366	0.21	0.63	239	-0.08	0.63	250	-0.22	0.66	0.46	0.67	0.22
Wave 2	322	0.73	0.53	218	0.47	0.73	216	0.42	0.78	0.41	0.47	0.07
Wave 3	302	0.97	0.46	188	0.91	1.07	185	0.87	1.06	0.08	0.13	0.04
Inhibition												
Wave 1	357	0.32	0.99	239	-0.14	0.74	248	-0.32	0.77	0.53	0.73	0.24
Wave 2	322	1.15	1.09	217	-0.01	0.71	216	-0.02	0.82	1.29	1.23	0.01
Wave 3	303	1.4	1.06	187	0.34	0.73	185	0.28	0.74	1.18	1.24	0.08

Age=age in months, STM-short-term memory, Exec-WM=executive component of working memory, Visual-WM=Visual-spatial working memory.

Table 3

Hierarchical Growth Modeling Predicting Math Calculation Performance

	Unconditional Means Model		Demographic Model		Executive Model 1		GF-Gc Model 2		STM Model 3		Full Model 4	
Fixed Effects	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	0.56***	0.02	0.44***	0.06	0.24***	0.05	0.26***	0.05	0.17***	0.05	0.25***	0.06
Age			0.01***	0.002	0.01***	0.002	0.01***	0.002	0.01***	0.002	0.01***	0.002
Age x cohort			0.01**	0.004	0.01*	0.004	0.01*	0.004	0.01	0.004	0.01	0.004
Gender			0.04	0.04	0.02	0.04	0.02	0.03	0.02	0.03	0.02	0.03
Hispanic			-0.17	0.10	-0.13	0.10	-0.14	0.08	-0.11	0.09	-0.13	0.08
Anglo			0.09	0.10	0.07	0.09	0.03	0.08	0.10	0.08	0.05	0.08
SES			0.09***	0.02	0.08**	0.02	0.02	0.02	0.07**	0.02	0.03	0.02
CELDT vs. no CELDT			0.10	0.12	0.07	0.08	0.08	0.05	0.08	0.09	0.07	0.08
Proficient. vs. less Proficient ELLs			0.20**	0.09	0.10	0.06	0.08*	0.04	0.08	0.06	0.07	0.06
Mono vs. Proficient ELLs			0.21	0.16	0.07	0.10	0.12	0.06	0.06	0.11	0.10	0.11
Read							0.15***	0.01			0.13***	0.01
Fluid							0.13***	0.02			0.13***	0.02
STM									0.05***	0.01	0.02	0.01
Speed									-0.12***	0.02	-0.07**	0.02
Exec-WM					0.09***	0.01	0.05***	0.01	0.08***	0.01	0.04**	0.01
Vis-WM					0.10***	0.02	0.07**	0.02	0.09***	0.02	0.07**	0.02
Inhibition					0.11***	0.02	0.09***	0.01	0.09***	0.02	0.08***	0.01

SE= standard error,*$p < .05$, ** $p < .01$, *** $p < .001$,

Gf-Gc=Fluid and Crystallized intelligence, Age=Chronological age centered at wave 1, CELDT= California English Language Development Test-, SES=Federal free lunch program, Fluid=Fluid intelligence, STM-short-term memory, Speed=Naming Speed. Exec-WM=executive component of working memory, Visual-WM=Visual-spatial working memory.

Table 4
Hierarchical Growth Modeling Predicting Math Calculation Performance

	Unconditional		Demographic		Model 1		Model 2		Model 3		Model 4	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Slope (linear)	0.13***	0.01	0.12***	0.01	0.14***	0.02	0.10***	0.02	0.13***	0.02	0.09***	0.02
Read							0.01***	0.0006			-0.006	0.005
Fluid							-0.001	0.001			0.02	0.01
STM									-0.01	0.01	-0.007	0.007
Speed									-0.04***	0.01	-0.04***	0.009
Exec-WM					0.03***	0.01	0.03***	0.01	0.03***	0.01	0.02**	0.008
Vis-WM					0.02	0.01	0.01	0.01	0.01	0.01	0.004	0.01
Inhibition					0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01
Slope (nonlinear)	-0.0004***	0.00005	-0.0003***	0.00005	-0.0005***	0.00008	-0.0003***	0.0001	-0.0005***	0.00008	-0.0002	0.0001
Read							0.00004	0.00003			0.00005	0.00003
Fluid							-0.00009	0.00006			-0.00007	0.00006
STM									0.00006*	0.00003	0.00004	0.00003
Speed									0.0002**	0.00005	0.0002***	0.00005
Exec-WM					-0.0001***	0.00003	-0.0001***	0.00003	-0.0001**	0.00004	-0.00009*	0.00004
Vis-WM					-0.00007	0.00006	-0.00006	0.00006	-0.00003	0.00006	-0.00002	0.00006
Inhibition					-0.00005	0.00005	-0.00006	0.00005	-0.00004	0.00005	-0.00005	0.00005
Random Effects												
τ_{01}	0.20***	0.02	0.16***	0.01	0.12***	0.01	0.08***	0.01	0.10***	0.01	0.08***	0.01
τ_{02}	0.006**	0.0005	0.004***	0.0006	0.003***	0.0004	0.002**	0.000336	0.003***	0.0004	0.002***	0.00032
τ_{03}	0.0003***	0.00005	0.0002***	0.00005	0.00006*	0.00003	0.00003	0.00003	0.00006*	0.00003	0.00003	0.00003
σ^2	0.22***	0.01	0.23***	0.01	0.23***	0.01	0.22***	0.01	0.22***	0.01	0.22	0.01
Fit Indices												
Deviance	4482.6		4287.10		3958.7		3653.1		3813.8		3570.1	
AIC	4496.6		4319.10		4008.7		3715.1		3875.8		3644.1	
BIC	4529.8		4394.20		4126.2		3860.3		4021.4		3817.4	

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

Note. SE= standard error, Fluid=Fluid intelligence, STM-short-term memory, Speed=Naming Speed. Exec-WM=executive component of working memory,

Vis-WM=Visual-spatial working memory. τ_{01} variation of intercepts of performance nested within classroom, τ_{02} covariance between intercept and slope, τ_{03} variation of student rate of change within classrooms, σ^2_0 reflects the residual variation of students within classroom, Deviance= Chi-square value for the correspondence between model and data, AIC= Akaike's Information Criterion, BIC Bayesian Criterion.

Figure 1

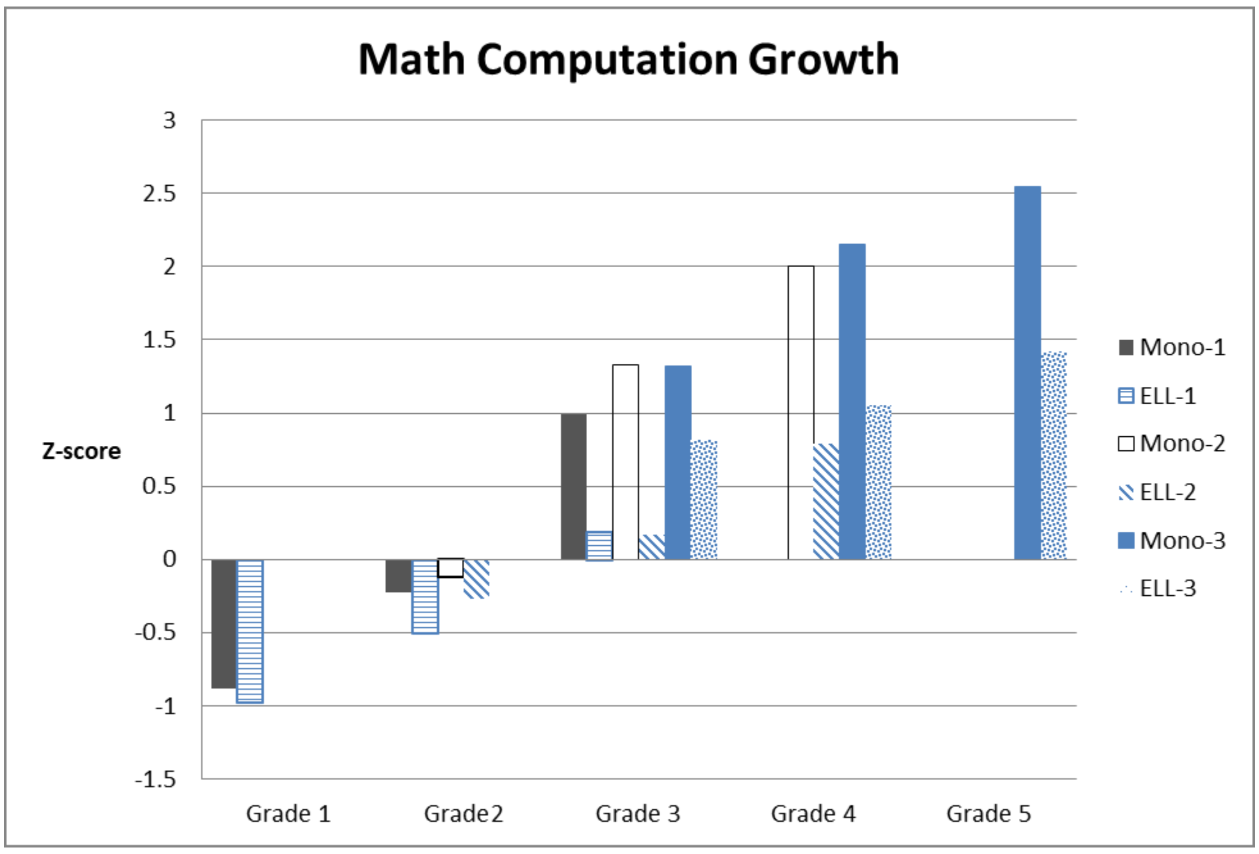

Figure 1. Mono=monolingual children, ELL=English language learners, -1=cohort 1 grades 1 to 3, -2=Cohort 2 grades 2 to 4, -3=Cohort 3 grades 3 to 5. Solid bars are monolingual children and bars with designs are ELL children.

Appendix A

Standardized Estimates for Total Sample Used to Create Latent Measures from Wave 1

Reading		Mean	SD	Reliability	Loading	SE	t-ratio
1	Word-ID	103.64	15.28	0.91	0.62	0.06	10.02***
2	Comprehension	96.68	15.16	0.88	0.99	0.08	13.62***
Naming Speed							
3	Numbers	45.25	16.98	0.88	0.91	0.09	9.76***
4	Letter	49.46	19.16	0.92	0.63	0.08	8.30***
Inhibition							
5	Letters	6.52	2.96	0.72	0.55	0.08	6.97***
6	Numbers	4.54	3.03	0.71	0.57	0.08	7.17***
Short-term Memory (phonological loop)							
7	Forward Digit	5.85	1.41	0.87	0.38	0.06	5.84**
8	Backward	2.63	1.06	0.85	0.73	0.04	18.81***
9	Word Span	3.36	1.45	0.84	0.83	0.03	25.62***
10	Nonword	3.48	2.29	0.86	0.55	0.05	10.09***
Visual-Spatial-WM							
12	Matrix	5.39	1.59	0.84	0.49	0.1	4.78**
13	Mapping	3.48	1.34	0.85	0.35	0.09	3.99**
Executive-WM							
14	Listening span	3.06	1.44	0.87	0.67	0.05	14.27***
15	Concept. Span	3.52	1.33	0.86	0.80	0.04	20.35***
16	Updating	4.20	4.22	0.87	0.55	0.06	9.74***

**p < .01, ***p < .001; Word ID=word identification, Comprehension=passage comprehension
Matrix=visual matrix, Mapping=mapping & direction. Concept=conceptual span