Contemporary Educational Psychology 60 (2020) 101825

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

# Language-related longitudinal predictors of arithmetic word problem solving: A structural equation modeling approach

Mercedes Spencer*, Lynn S. Fuchs, Douglas Fuchs

*Vanderbilt University, USA*

ARTICLE INFO

*Keywords:*
Mathematics
Structural equation modeling
Multigroup modeling
Word problems
Longitudinal

ABSTRACT

We investigated the longitudinal relations between cognitive skills, specifically language-related skills, and word-problem solving in 340 children (6.10–9.02 years). We used structural equation modeling to examine whether word-problem solving, computation skill, working memory, nonverbal reasoning, oral language, and word reading fluency measured at second grade were associated with performance on measures of word-problem solving in fourth grade. Results indicated that prior word-problem solving, computation skill, nonverbal reasoning, and oral language were significantly associated with children's later word-problem solving. Multi-group modeling suggested that these relations were not significantly different for boys versus girls. Implications of these findings are discussed.

## 1. Introduction

By the end of fourth grade, only 40% of students in the United States are scoring at or above proficiency in mathematics (National Center for Education Statistics, 2017). This is problematic because early mathematics skills are associated with academic and life outcomes (Claessens, Duncan, & Engel, 2009; Magnuson, Duncan, Lee, & Metzger, 2016; Rivera-Batiz, 1992). Given this, we investigated the longitudinal relations between early cognitive skills and later arithmetic word-problem solving. We focused on word-problem solving because the ability to solve word problems is significantly associated with later life outcomes (Murnane, Willett, Braatz, & Duhaldeborde, 2001) and are emphasized within key educational initiatives (e.g., Common Core State Standards, the National Mathematics Advisory Panel, and statewide high-stakes assessments; Cohen, Gregg, & Deng, 2005; Common Core State Standards Initiative, 2010; National Mathematics Advisory Panel, 2008).

Word problems are "verbal descriptions of problem situations wherein one or more questions … can be [answered] … by the application of mathematical operations to numerical data available in the problem statement" (Verschaffel, Greer, & De Corte, 2000, p. ix). For example, "John has nine balloons. He gave away three balloons to his friend, Mark. How many balloons does John have now?" In this instance, the problem solver needs to identify relevant linguistic and numerical information (nine and three) and the appropriate operands indicated in the narrative (*gave away*) to solve the problem (six balloons). Thus, it is not surprising that word-problem solving requires the consolidation of a variety of cognitive skills, including non-mathematical processes such as working memory, reasoning ability, and oral language (Andersson, 2007; Fuchs et al., 2006; Fuchs, Fuchs et al., 2008; Jordan, Levine, & Huttenlocher, 1995).

The complexity of word problems, and subsequent recruitment of multiple cognitive processes, may explain why word problems tend to be more difficult for children to solve than de-contextualized numerical problems (Cummins, Kintsch, Reusser, & Weimer, 1988; Daroczy, Wolska, Meurers, & Nuerk, 2015). That is, solving the mathematical problem 9 – 3 = __ (the numerical equivalent of the above word problem) is easier than a word problem including the same numerical information but presented verbally. This increased difficulty may explain why some children who struggle to solve word problems perform adequately on other measures of mathematic skill (e.g., Swanson, Jerman, & Zheng, 2008). Further, although arithmetic skills are foundational to word-problem solving, a unique set of cognitive skills is associated with word-problem solving separate from arithmetic or calculations (Fuchs et al., 2006; Fuchs, Zumeta et al., 2010; Swanson & Beebe-Frankenberger, 2004).

Because word-problem solving ability appears to be a discrete skill (e.g., Fuchs et al., 2006) strongly associated with mathematical competence (Cohen et al., 2005) and life outcomes (Murnane et al., 2001), we aimed to investigate the association between early (second grade) word-problem solving, arithmetic skill, nonverbal reasoning, working memory, oral language, and word reading fluency and later (fourth grade) word-problem solving. Below, we review several relevant theoretical frameworks of word-problem solving. Next, we provide empirical and theoretical rationale for the inclusion of the chosen cognitive variables. Following this, we discuss the importance of examining sex differences in word-problem solving. Finally, we provide a brief description of prior studies focused on sex differences in word-problem solving and a rationale for the present study.

* Corresponding author at: The Department of Special Education, Vanderbilt University, PMB 40, 230 Appleton Place, Nashville, TN 37203, USA.
*E-mail addresses:* mercedes.spencer@vanderbilt.edu (M. Spencer), lynn.fuchs@vanderbilt.edu (L.S. Fuchs), doug.fuchs@vanderbilt.edu (D. Fuchs).

https://doi.org/10.1016/j.cedpsych.2019.101825

Available online 26 November 2019
0361-476X/ 

### 1.1. Theoretical frameworks of word-problem solving

Given the multifaceted nature of the word-problem solving process, several theoretical frameworks have been proposed. In their comprehensive report, the Mathematics Learning Study Committee and National Research Council (NRC; Mathematics Learning Study Committee & National Research Council, 2001) outlined five foundational skills, referred to a strands, thought to be critical to the development of mathematics proficiency. These include mathematical knowledge (conceptual understanding) as well as the ability to carry out mathematical operations (procedural fluency), apply appropriate problem-solving strategies (strategic competence), and think critically and logically (adaptive reasoning); individuals additionally need to believe that mathematics is a useful skill (productive disposition; see also Mayer, 2013).

Although this framework describes the development of mathematical proficiency in broad terms, each of these skills are important for word-problem solving competency more specifically. For instance, word problems often include descriptions of real-world situations (Staub & Reusser, 1995; productive disposition) and require an individual to understand the nature of the word problem (mathematical knowledge; Seh Bae, Chiang, & Hickson, 2015), apply necessary computational skills (procedural fluency; Muth, 1984) and strategies (strategic competence; De Corte & Verschaffel, 1987), and make logical inferences (adaptive reasoning; Tajika, Nakatsu, Nozaki, Neumann, & Maruno, 2007) in order to successfully derive a solution. Several of these strands are also embedded within word-problem-specific theoretical frameworks, which we describe in greater detail below.

Kintsch and Greeno's (1985) dual representation model asserts that, given the format of word problems, an individual's text comprehension skill impacts the ability to solve word problems. Further, individuals' experience(s) with similar texts and information allows them to build a situation model that subsequently facilitates comprehension of the text (see Van Dijk & Kintsch, 1983) and likely aids word-problem solving. Therefore, linguistic knowledge is identified as a key component of comprehending the text and subsequently solving the word problem; exposure to the text provides an opportunity for the individual to create a mental representation of the information (situation model) and merge this with prior knowledge to identify a problem model (Kintsch & Greeno, 1985).

Thus, successful word-problem solvers integrate their knowledge of mathematical problem solving (e.g., rules related to sums and minuends and conceptual understanding; Mathematics Learning Study Committee, National Research Council, 2001) and corresponding verbal information contained in the problem statement (e.g., activating verbal knowledge for words/phases such as *have altogether* and *less than* and their corresponding operands and procedural fluency; Mathematics Learning Study Committee, National Research Council, 2001; Powell, Fuchs, Fuchs, Cirino, & Fletcher, 2009; Thevenot, Devidal, Barrouillet, & Fayol, 2007). As such, word-problem solving additionally relies on cognitive processes beyond those required for text comprehension (e.g., word-problem-specific language comprehension; Fuchs, Fuchs, Compton, Hamlett, & Wang, 2015).

Hegarty, Mayer, and Monk (1995) also proposed a more linguistic-based model. In their model, reading the word problem prompts an individual to create or update their semantic situation model via a direct translation approach (i.e., identify and integrate some semantic and corresponding mathematical information) or the problem model approach (i.e., create or update the problem or situation model to adequately fit with the word problem as opposed to relying on a subset of features). With either approach, the individual then creates a plan to solve the word problem. Individuals who take the direct route are more likely to encounter word-problem solving difficulties because they have an incomplete situation model due to their reliance on a subset of the words/numbers. Individuals who take the problem model approach experience greater success because they have a more comprehensive understanding of the problem (Hegarty et al., 1995). These frameworks are supported by studies showing that word problems with simpler language are easier to comprehend and solve than those with more complex language (e.g., Cummins et al., 1988) and that word-problem features, such as words or operands, affect word-problem solution accuracy (Daroczy et al., 2015; Koedinger & Nathan, 2004).

### 1.2. Cognitive skills associated with word-problem solving

Previous research has identified several relatively robust cognitive skills that are associated with word-problem solving, including arithmetic competence, nonverbal reasoning, working memory, oral language, and word reading fluency. Based on the previously discussed theoretical frameworks of word-problem solving, arithmetic competence and nonverbal reasoning would facilitate the development of situation and problem models (e.g., Kintsch & Greeno, 1985; Hegarty et al., 1995) and the application of logical inference-making and strategies (Mathematics Learning Study Committee, National Research Council, 2001). Working memory would facilitate the activation of situation models, text comprehension, and formation of the problem model, while enabling individuals to hold and manipulate representations in memory. Oral language would facilitate an understanding of verbal information in the word problem and the creation a situation and problem model. We next discuss empirical evidence for the relations between word-problem solving and each of these cognitive skills.

#### 1.2.1. Arithmetic skill

Arithmetic skill, defined as the ability to solve single-digit calculation problems (Fuchs et al., 2006), is distinct from word-problem solving skill (Fuchs, Fuchs et al., 2008; Seethaler, Fuchs, Fuchs, & Compton, 2012). However, arithmetic is robustly associated with word-problem solving (Fuchs, Fuchs et al., 2008; Kail & Hall, 1999; Wang, Fuchs, & Fuchs, 2016). For example, children with mathematical disabilities (MD) tend to perform worse on word-problem solving than children without MD (e.g., Fuchs & Fuchs, 2002). These relations likely exist because word problems require the application of mathematical knowledge (Verschaffel et al., 2000) while also tapping pre-algebraic knowledge in some instances (Fuchs, Zumeta et al., 2010).

#### 1.2.2. Nonverbal reasoning

Nonverbal reasoning, the ability to engage in visual problem solving (Kroger et al., 2002), is associated with word-problem solving both concurrently and longitudinally (Fuchs et al., 2005, 2006; Tolar et al., 2012). Reasoning skills are often uniquely related to word-problem solving over and above other well-established cognitive skills (e.g., number line estimation, number sets, language, attention, working memory, and processing speed; Fuchs, Geary et al. 2010), likely because mathematical reasoning underlies the development of the problem model (Quilici & Mayer, 1996). Reasoning may also affect an individual's application of strategies and the understanding of whether a correct solution is obtained (e.g., Xin, Jitendra, & Deatline-Buchman, 2005).

#### 1.2.3. Working memory

Working memory, or the ability to maintain and process information (Baddeley, 1992), is related to word-problem solving outcomes (e.g., Andersson, 2007; Cummins et al., 1988; Passolunghi & Siegel, 2001; Zheng, Swanson, & Marcoulides, 2011). For instance, children with problem solving difficulties often have associated weaknesses in working memory (e.g., Passolunghi & Mammarella, 2010; Swanson & Beebe-Frankenberger, 2004). Working memory also often explains variance in word-problem solving over and above a variety of other cognitive skills (e.g., phonological processing, IQ, reading comprehension, and calculation skill; Swanson, 2004).

#### 1.2.4. Oral language

Oral language is a broad term used to describe various language-based skills, including vocabulary knowledge, listening comprehension, and syntactic knowledge among others. Numerous studies demonstrate that oral language plays a strong role in both text and language comprehension (e.g., Hoover & Gough, 1990; Kendeou, Van den Broek, White, & Lynch, 2009; Marchman & Fernald, 2008; Roth, Speece, & Cooper, 2002) and that these same variables are related to word-problem solving (Fuchs et al., 2006, 2015, 2018; Kintsch & Greeno, 1985;

Pina, Fuentes, Castillo, & Diamantopoulou, 2014; Wang et al., 2016). In fact, children with word-problem solving difficulties tend to exhibit corresponding language deficits (e.g., Fuchs, Fuchs et al., 2008). This association may occur because language deficits are often associated with general arithmetic difficulties (Donlan, Cowan, Newton, & Lloyd, 2007) and because oral language potentially facilitates development of mathematical skill (Fletcher, Lyon, Fuchs, & Barnes, 2007; Purpura, Hume, Sims, & Lonigan, 2011).

#### 1.2.5. Word reading fluency

Word reading fluency, or the ability to read words quickly and accurately, is associated with both text comprehension (Klauda & Guthrie, 2008; Petscher & Kim, 2011) and word-problem solving (Fuchs et al., 2006). Children with poor reading fluency often have word-problem solving deficits. For instance, Fuchs and Fuchs (2002) examined word-problem solving performance for children categorized as having MD (scoring below 1.5 standard deviations [*SD*] on computational fluency), combined MD and reading disability (scoring below 1.5 *SD* on computational fluency and reading fluency), and as being typically developing. Children with MD and reading disability performed lower on complex word problems and real-world problem solving relative to typically developing children and children with MD only.

### 1.3. Sex differences

Although research suggests that the prevalence of MD is similar across males and females (Lewis & Fisher, 2016), empirical evidence indicates that males tend to outperform females on arithmetic and word-problem solving (e.g., Hedges & Nowell, 1995; Jõgi & Kikas, 2016; Low & Over, 1993). However, findings on sex differences in math are often mixed (see for example, Vilenius-Tuohimaa, Aunola, & Nurmi, 2008). This may occur because word-problem solving is potentially affected by a variety of other factors, such as language competencies that tend to favor females (e.g., Cole, 1997; Loveless, 2015; Voyer & Voyer, 2014), psychological characteristics, such as math anxiety and/or strategy use (e.g., Davis & Carr, 2002; Devine, Fawcett, Szűcs, & Dowker, 2012), and methodological features of math assessments (e.g., Berberoglu, 1995). Therefore, it remains important to investigate whether relations between cognitive correlates of word-problem solving skills differ for boys versus girls.

### 1.4. The present study

The aim of present study was to extend previous findings in several important ways. First, although multiple studies have examined sex differences in mathematical performance (e.g., Carr & Jessup, 1997; Geary, Saults, Liu, & Hoard, 2000; Marshall & Smith, 1987; Royer, Tronsky, Chan, Jackson, & Marchant, 1999) and word-problem solving (Chipman, Marshall, & Scott, 1991; Quinn & Spencer, 2001; Walsh, Hickey, & Duffy, 1999), few have examined sex differences in longitudinal word-problem-solving outcomes (see Jõgi & Kikas, 2016; Björn, Aunola, & Nurmi, 2016). Third, despite prior investigations of associations between cognitive skills and word-problem solving, several relied on single-time point data (e.g., Kyttälä & Björn, 2014), included sex as the independent variable (e.g., Jõgi & Kikas, 2016; Vilenius-Tuohimaa et al., 2008), or did not examine sex differences at all (e.g., Fuchs et al., 2015; Wang et al., 2016).

Jõgi and Kikas (2016) examined word-problem solving outcomes in 864 children from first to third grade who were assessed on calculation skill, word-problem solving, nonverbal IQ, language, executive function, and task persistence assessed in first and third grades. All variables, including sex, explained unique variance in third-grade word-problem solving. Björn et al. (2016) examined longitudinal relations among fluency, reading comprehension, calculations, and word-problem solving in a sample of 224 fourth graders, who were also assessed in seventh and ninth grade. Fourth-grade reading comprehension was associated with word-problem solving in in seventh and ninth grades for boys and girls, respectively. Similarly, in a sample of 99 eighth graders, Kyttälä and Björn (2014) found that reading comprehension was significantly associated with word-problem solving for boys while technical reading was related to word-problem solving for girls and boys. None of these previous studies, however, examined the contributions of prior word-problem solving, reasoning skills, working memory, oral language, and word reading fluency within a single model while considering sex differences. Thus, previous findings are capable of identifying whether boys and girls may perform differently on a specific outcome measure but reveal little about how individual differences in the development of word-problem solving occur and potential sex differences in relations between other variables of interest. Further, in the few instances when researchers have examined boys and girls separately, those studies did not include measures of oral language or working memory in their models (e.g., Björn et al., 2016; Kyttälä & Björn, 2014).

We therefore extended previous work by bringing together the methods of (a) studies that did not examine sex differences but included a rich cognitive battery (e.g., Fuchs et al., 2015; Wang et al., 2016) with (b) studies that examined sex differences but with a more limited set of assessments. We also sought to extend previous findings on the relations between cognitive processes and word-problem solving (Andersson, 2007; Fuchs, Fuchs et al., 2008; Swanson & Beebe-Frankenberger, 2004; Tolar et al., 2012) by examining associative relations across second to fourth grade, when children are building foundational skills related to word-problem solving competency (Wilson, 2009) and by considering how findings support theoretical models of word-problem processing.

We were also interested in the whether our findings would support text comprehension models of word-problem solving (e.g., dual representation model; Kintsch & Greeno, 1985) by examining the extent to which foundation reading comprehension-related skills (word reading and oral language; Hoover & Gough, 1990) were also related to word-problem solving longitudinally. Further, we aimed to extend the findings of three studies by Fuchs and colleagues who across studies, investigated the longitudinal cognitive correlates of text comprehension versus word-problem solving (Fuchs et al., 2015, 2018) and pre-algebraic knowledge versus word-problem solving (Fuchs et al., 2016). The present study builds on these previous studies by examining similar variables over a wider developmental span (Fuchs et al., 2015, 2018) and assessing whether the pattern of effects is similar for boys and girls within a latent variable framework (Fuchs et al., 2016). Thus, findings of the present study add in important ways to the present literature and to theoretical frameworks of word-problem solving.

## 2. Materials and methods

Participants were four cohorts of second graders, ages 6.10 to 9.02 years ($M = 7.63$, $SD = 0.43$) near the start of the study, who were part of a larger longitudinal intervention study (Fuchs et al., 2014) conducted in the Southeastern United States. Study approval was granted by the Vanderbilt University Institutional Review Board prior to conducting the study (IRB # 130999). In the present analysis, we only included children who did not receive intervention as part of the parent study, because intervention was designed to alter the natural course of development. The sample included 341 children from 39 classrooms. Participants were identified as 41.3% African American, 25.8% white non-Hispanic, 25.8% white Hispanic, 2.9% Kurdish, and 4.2% other (9.1% did not specify). The sample was 56.5% female (9.1% did not specify).

### 2.1. Measures

#### 2.1.1. Start-of-second-grade variables

The variables in our models were word-problem solving, computation skill, nonverbal reasoning, working memory, oral language, and word reading fluency (see Table 1).

With *Story Problems* (Jordan & Hanich, 2000), students solve word problems with sums and minuends less than 10. This assessment contains four word-problem types (change, combine, compare, equalize) with a total of 14 items. Items are read aloud, while students see the

**Table 1**
Measures in the Current Study.

| Measure | Skill(s) Assessed | Assessment Features |
|---|---|---|
| Story Problems[a] | Word-problem solving | Solve up to 14 orally-presented word problems (change, combine, compare, equalize) with sums less than 10 |
| Addition 0–12[a] | Arithmetic fluency | Solve up to 25 single-digit addition problems with sums ranging from 6 to 12 |
| Addition 0–18[a] | Arithmetic fluency | Solve up to 25 single-digit addition problems with sums ranging from 5 to 18 |
| DD Addition[a] | Calculation fluency | Solve up to 20 double-digit addition problems with and without regrouping |
| Subtraction 0–12[a] | Arithmetic fluency | Solve up to 25 single-digit subtraction problems with minuends ranging from 6 to 12 |
| Subtraction 0–18[a] | Arithmetic fluency | Solve up to 25 single-digit subtraction problems with minuends ranging from 5 to 18 |
| DD Subtraction[a] | Calculation fluency | Solve up to 20 double-digit subtraction problems with and without regrouping |
| WASI MR | Nonverbal reasoning | Complete a missing portion of a matrix; multiple-choice |
| WJ-III CF | Rule/concept understanding | Identify rules for concepts based on previously presented exemplars |
| WMTB-C LR | Working memory | Recall the last word for up to 6 orally-presented sentences across multiple spans |
| WMTB-C CR | Working memory | Recall numerals of counted sets for up to 6 visually-presented cards across multiple spans |
| WASI Vocabulary | Vocabulary | Provide definitions for up to 42 presented pictures/orally-presented words |
| WDRB LC | Language comprehension | Provide a missing word for up to 38 orally-presented sentence/passages |
| WIF-B [a] | Word reading fluency | Read a list of words |
| WIF-N [a] | Word reading fluency | Read a list of words |
| ITBS9 [b] | Word-problem solving | Solve up to 24 orally-presented word problems based on information presented in graphs/tables; multiple-choice |
| ITBS10 [b] | Word-problem solving | Solve up to 24 orally-presented word problems based on information presented in graphs/tables; multiple-choice |
| VSP [b] | Word-problem solving | Solve up to 18 orally-presented word problems (total, difference, change) both with and without irrelevant information and with and without graphs |

[a] Timed.
[b] Outcome; DD = Double digit; WASI = Wechsler Abbreviated Scale of Intelligence; MR = Matrix Reasoning; WJ = III = Woodcock-Johnson, Third Edition; CF = Concept Formation; WMTB-C = Working Memory Test Battery for Children; LR = Listening recall; CR = Counting recall; WDRB = Woodcock Diagnostic Reading Battery; LC = Listening comprehension; WIF-B/N = Word identification fluency form B/N; VSP = Vanderbilt Story Problems; ITBS9/10 = Iowa Test of Basic Skills level 9/10.

text; they have 30 sec to respond. Answers are scored for correct numerical answers. On the full sample, α was 0.83 to 0.92.

Four subtests from *Math Fact Fluency* (Fuchs, Hamlett, & Powell, 2003) and two subtests from the *Double-Digit Addition* and *Subtraction Tests* (Fuchs et al., 2003) were used to assess addition and subtraction fluency in second grade. With *Math Fact Fluency*, students solve single-digit addition and subtraction problems with sums and minuends from 6 to 12 for the first group of items and 5 to 18 for the second group of items (up to 25 items per subtest). Students have 1 min for each subtest. Test-retest reliability is 0.87 (Fuchs et al., 2016). With *Double-Digit Addition* and *Subtraction*, students solve double-digit addition and subtraction problems (up to 20 items in each subtest) within 5 min per subtest. As per Fuchs et al. (2003) and Fuchs and Seethaler (2008), α = 0.94 and 0.92 for Double-Digit Addition and Subtraction, respectively.

With *Wechsler Abbreviated Scale of Intelligence* (WASI; Wechsler, 1999) – *Matrix Reasoning*, students are presented with a matrix that has a portion missing and are given five options to complete the matrix. Answers are generated verbally or by pointing to an option. Reliability for this subtest is 0.94 (Wechsler, 1999).

With *Woodcock-Johnson Tests of Achievement-Concept Formation (3rd ed.* WJ-III; (Woodcock, McGrew, & Mather, 2001), students identify rules for concepts based on previously presented exemplars. Reliability for this assessment is 0.94 (Schrank, McGrew, & Woodcock, 2001).

With *Working Memory Test Battery for Children* (WMTB-C; Pickering & Gathercole, 2001) – *Counting Recall* and *Listening Recall*, students recall a series of items. Each subtest has six items at span levels 1–6 to 1–9. Passing four items at a level moves the child to the next, increasing the number of items to be remembered by one. Failing three items within a level terminates the subtest. The score is trials correct. For *Counting Recall* (working memory-numerals), children count 4, 5, 6, or 7 dots on a series of cards; then they recall the numerals of the counted sets. At second grade, subtest stability is 0.83–0.85. For *Listening Recall* (working memory-words), children determine if each sentence in a series is true; then they recall the last word of each sentence.

With WASI (Wechsler, 1999) – *Vocabulary*, the tester presents up to 42 pictures or oral words; students provide definitions. Answers are awarded a score of 0, 1 or 2 depending on the definition's quality. Testing stops after five consecutive errors. Split-half reliability is 0.86–0.87 (Zhu, 1999).

With *Woodcock Diagnostic Reading Battery – Listening Comprehension* (WDRB; Woodcock, 1997), the tester presents up to 38 sentences or passages of varying difficulty (e.g., verbal analogies, discerning implications); for each, students restore a missing word. Testing stops after six consecutive errors. Reliability is 0.80 (Woodcock, 1997).

With *Word Identification Fluency* (WIF; Fuchs, Fuchs, & Compton, 2004), students have 1 min to read a list of words. When students hesitate more than three sec, the tester moves them to the next word. The score is the average words read across two lists. Alternate-form reliability is 0.88 to 0.97 (Fuchs et al., 2004).

### 2.1.2. Fourth-grade outcome measures

Word-problem solving was measured using *Vanderbilt Story Problems* (VSP; Fuchs & Seethaler, 2008) and two versions (Levels 9 and 10) of the Iowa Test of Basic Skills (ITBS; Hoover, Hieronymus, Frisbie, & Dunbar, 1993; see Table 1). Items are read aloud; students see the text. VSP measures the ability to solve combine, compare, and change word problems of varying degrees of complexity (with and without irrelevant information; with and without graphs) across 18 items. One point is awarded each for math answer (i.e., arithmetic computation) and one point for the label (i.e., word-problem processing). The score is the sum of correct math answers and labels. As per Fuchs et al. (2009), α = 0.86.

With ITBS, students solve up to 24 multiple-choice word problems of varying difficulty (e.g., single- and multi-step calculation; determining sums and differences). Items are read aloud, while students see the text. Kuder-Richardson 20 internal consistency values range from 0.84 to 0.92 for the mathematics portion of the ITBS Levels 9 and 10 (Dunbar et al., 2015).

## 2.2. Procedure

Students were assessed in September-October of second grade. Math Fact Fluency and Double-Digit Addition and Subtraction were administered in whole-class format; WASI, WMTB-C, WJ-III, WDRB, and WIF were administered individually. Students were assessed on the outcomes in April of fourth grade in small-group format. All directions were provided orally. Trained research assistants administered all assessments, and data collection was standardized using written directions. Individual assessments were audio recorded, with 20% scored for procedural accuracy (greater than 97%).

## 2.3. Analyses

We used confirmatory factor analysis (CFA) modeling and structural

**Table 2**
Descriptive Statistics Across the Full Sample.

| Measure | *N* | Min. | Max. | Mean | *SD* | Skew | Kurtosis |
|---|---|---|---|---|---|---|---|
| Age | 338 | 6.10 | 9.02 | 7.630 | 0.434 | 0.231 | 0.522 |
| Sex | 309 | 1 | 2 | 1.437 | 0.497 | 0.256 | −1.947 |
| Story Problems[a] | 339 | 0 | 14 | 6.501 | 3.417 | 0.298 | −0.697 |
| Addition 0–12[a] | 338 | 1 | 22 | 8.411 | 4.644 | 0.738 | 0.067 |
| Addition 0–18[a] | 338 | 0 | 16 | 6.056 | 3.481 | 0.534 | −0.066 |
| DD Addition[a] | 335 | 0 | 12 | 3.260 | 3.049 | 0.771 | −0.232 |
| Subtraction 0–12[a] | 338 | 0 | 12 | 4.320 | 2.608 | 0.665 | 0.200 |
| Subtraction 0–18[a] | 338 | 0 | 11 | 3.033 | 2.375 | 0.925 | 0.763 |
| DD Subtraction[a] | 335 | 0 | 9 | 2.290 | 2.659 | 0.985 | −0.042 |
| WASI –MR | 340 | 2 | 28 | 11.374 | 5.771 | 0.654 | −0.546 |
| WJ-III CF | 335 | 1 | 27 | 11.272 | 5.589 | 0.577 | −0.019 |
| WMTB-C LR | 338 | 0 | 16 | 7.092 | 3.766 | −0.091 | −0.420 |
| WMTB-C CR | 338 | 2 | 26 | 14.192 | 4.436 | 0.104 | −0.146 |
| WASI Vocabulary | 340 | 5 | 37 | 21.853 | 6.535 | −0.010 | −0.049 |
| WDRB LC | 340 | 5 | 29 | 16.500 | 4.715 | −0.401 | −0.076 |
| WIF-B | 338 | 0 | 111 | 41.794 | 24.811 | 0.382 | −0.630 |
| WIF-N | 337 | 1 | 120 | 55.941 | 22.377 | −0.116 | −0.319 |
| ITBS9[b] | 283 | 1 | 17 | 10.131 | 3.684 | −0.193 | −0.621 |
| ITBS10[b] | 283 | 1 | 11 | 5.452 | 2.167 | 0.334 | −0.445 |
| VSP[b] | 283 | 0 | 29 | 11.004 | 6.627 | 0.596 | 0.021 |

*Note.* Raw scores reported.
[a] Timed.
[b] Outcome; *SD* = Standard deviation; Min. = Minimum; Max. = Maximum; Age = Age in second grade; DD = Double digit; WASI = Wechsler Abbreviated Scale of Intelligence; MR = Matrix Reasoning; WJ-III = Woodcock-Johnson, Third Edition; CF = Concept Formation; WMTB-C = Working Memory Test Battery for Children; LR = Listening recall; CR = Counting recall; Vocab. = Vocabulary; WDRB = Woodcock Diagnostic Reading Battery; LC = Listening comprehension; WIF-B/N = Word identification fluency form B/N; VSP = Vanderbilt Story Problems; ITBS9/10 = Iowa Test of Basic Skills level 9/10.

equation modeling (SEM) to examine the associations between start-of-second-grade variables and fourth-grade outcomes. We began by specifying CFA models in order to create latent variables for all proposed constructs. Next, we specified several SEMs that included paths from age and start-of-second-grade cognitive skills to fourth-grade word-problem solving outcomes. Across models, we included all variables and allowed the second-grade variables to correlate given the likelihood of anticipated associations across measures. We used a step-by-step model-testing approach to systematically remove paths (i.e., fixed these paths to zero within the model) that were nonsignificant (aside from age, which we retained as a control variable across models). We then compared these nested models using chi-square difference testing in order to identify the best-fitting and most parsimonious model (e.g., Bentler & Mooijaart, 1989; Mulaik et al., 1989). Across CFA and SEM models, we began by fitting models to the full sample followed by modeling girls and boys separately. Following this, we ran multi-group models to identify whether relations among skills differed across boys and girls.

Model fit was determined using several common indices, including the chi-square statistic ($\chi^2$), comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). CFI and TLI values greater than 0.95, RMSEA values at or below 0.08, and SRMR values at or below 0.05 are indicative of good model fit (Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996); CFI and TLI values greater than 0.90, RMSEA values at or below 0.10, and SRMR values at or below 0.08 are indicative of adequate model fit (Hu & Bentler, 1999; Kline, 2015; MacCallum et al., 1996).

## 3. Results

### 3.1. Preliminary analyses

Data were analyzed using SPSS software (Version 25.0; IBM Corp., 2017) and Mplus software (Version 7.11; Muthén & Muthén, 2013). Prior to SEM, we examined the full sample data for univariate and bivariate outliers, skewness and kurtosis, and missingness. Outliers were identified using the median plus or minus two interquartile ranges criterion. Based on this criterion, outliers accounted for 0.78% of the total data. The addition 0–18 subtest of the Math Fact Fluency also exhibited a kurtosis value (2.12) outside the acceptable range of plus or minus 2. Given these patterns in the data, we replaced outlier values with values at the highest or lowest end of the outlier range. We then identified one multivariate outlier using Mahalanobis distance across the 18 measures of interest ($p < .001$); this participant was removed from all subsequent analyses.

All variables demonstrated adequate variability and had skewness and kurtosis values within the acceptable range of plus or minus 2 (see Table 2 for descriptive statistics for the full sample; see Table A.1 in the appendix for descriptive statistics for boys and girls separately). We used Little's MCAR test to determine if the data were missing completely at random. Based on expectation-maximization statistics (means and correlations) for the measures of interest, including age and gender (as a categorical variable), the data were missing completely at random, $\chi^2 = 150.506$ (145), $p = .360$ (see Table 3 for correlations for the full sample; see Table B.1 in the appendix for correlations for boys and girls separately). Thus, we used maximum likelihood estimation in all subsequent SEMs. Students whose teachers did not specify the child's sex ($N = 31$) were excluded from single- (i.e., boys only and girls only models) and multi-group analyses examining sex differences.

### 3.2. CFA and SEM

For the ***full sample,*** we initially specified a CFA model that included all proposed constructs as latent variables (see Table C.1 in the appendix for factor loadings for all CFA models). Further, students' single- and double-digit addition and subtraction were specified as comprising a higher-order latent factor of Computation Skill (see Fig. 1). The variance of this factor was fixed to one in order to obtain factor loadings for all three factors. This model provided an excellent fit to the data (Model F1), suggesting that all the observed indicators were adequately represented by their respective latent variables. Initially, we specified the observed indicators of WASI Matrix Reasoning and WJ-III Concept Formation as comprising a single latent factor (Reasoning); however, there was a linear dependency between this factor and the fourth-grade Word-Problem Solving factor for the full sample ($r = 0.969$) and for boys ($r = 1.015$); correlations between these factors were strong for girls as well ($r = 0.928$). Therefore, these variables were instead included as observed variables across all models for consistency.

In our initial SEM (Model F2), we specified associations between all variables (age, start-of-second-grade word-problem solving, computation skill, matrix reasoning, concept formation, working memory, oral language, and word reading fluency) and fourth-grade word-problem solving (see Fig. 2a). This model provided an excellent fit to the data (see Table 4). The weakest parameters included working memory ($\beta = 0.062$, $p = .518$ in Model F2), word reading fluency ($\beta = 0.055$, $p = .375$ in Model F3), and concept formation ($\beta = 0.056$, $p = .279$ in Model F4). Each of these were paths were removed when specifying Models F3, F4, and F5, respectively (see Table D.1 in the appendix for parameter estimates across Models F2, F3, and F4). Chi-square difference testing indicated nonsignificant differences between Models F2 and F3, Models F3 and F4, and Models F5 and F4, suggesting that the most parsimonious model (Model F5) was the preferred model (see Table 4). This final model provided an excellent fit to the data ($\chi^2 = 162.537$ (120), $p = .006$, CFI = 0.985, TLI = 0.979, RMSEA = 0.032, 90% confidence interval [0.018, 0.044], p-close = 0.994; SRMR = 0.038). In this final model, students' age ($\beta = -0.107$, $p < .05$) and prior performance on oral language ($\beta = 0.351$, $p < .001$), word-problem solving ($\beta = 0.304$, $p < .001$), computation skill ($\beta = 0.386$, $p < .001$), and matrix reasoning ($\beta = 0.276$, $p < .001$) were all significantly associated with word-problem solving in fourth grade (see Table 5 and Fig. 2b). These variables accounted for 85.1% variance in children's fourth-grade word-problem solving.

**Table 3**
Correlations Across the Full Sample.

| Variables | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Age | – | | | | | | | | | | | | |
| 2. Sex | 0.110 | – | | | | | | | | | | | |
| 3. Story Problems | 0.030 | 0.126* | – | | | | | | | | | | |
| 4. Addition 0–12 | 0.045 | 0.043 | 0.362** | – | | | | | | | | | |
| 5. Addition 0–18 | 0.018 | 0.029 | 0.376** | 0.760** | – | | | | | | | | |
| 6. DD Addition | −0.011 | −0.007 | 0.277** | 0.527** | 0.479** | – | | | | | | | |
| 7. Subtraction 0–12 | 0.074 | 0.129* | 0.384** | 0.561** | 0.531** | 0.336** | – | | | | | | |
| 8. Subtraction 0–18 | 0.083 | 0.100 | 0.343** | 0.520** | 0.461** | 0.327** | 0.750** | – | | | | | |
| 9. DD Subtraction | −0.054 | −0.058 | 0.239** | 0.460** | 0.420** | 0.718** | 0.261** | 0.305** | – | | | | |
| 10.WASI MR | −0.025 | 0.063 | 0.278** | 0.178** | 0.249** | 0.076 | 0.187** | 0.160** | 0.104 | – | | | |
| 11. WJ-III CF | −0.011 | 0.020 | 0.471** | 0.239** | 0.266** | 0.158** | 0.232** | 0.227** | 0.129* | 0.326** | – | | |
| 12. WMTB-C LR | −0.004 | 0.012 | 0.392** | 0.187** | 0.147** | 0.138* | 0.147** | 0.140** | 0.052 | 0.285** | 0.309** | – | |
| 13. WMTB-C CR | 0.020 | 0.040 | 0.282** | 0.271** | 0.216** | 0.274** | 0.214** | 0.224** | 0.229** | 0.272** | 0.260** | 0.391** | – |
| 14. WASI Vocabulary | 0.015 | −0.018 | 0.368** | 0.084 | 0.076 | −0.007 | −0.027 | −0.031 | −0.059 | 0.175** | 0.377** | 0.321** | 0.198** |
| 15. WDRB LC | 0.029 | 0.089 | 0.401** | 0.140* | 0.140** | 0.040 | 0.132* | 0.131* | −0.005 | 0.183** | 0.393** | 0.412** | 0.206** |
| 16. WIF-B | −0.052 | 0.005 | 0.425** | 0.404** | 0.369** | 0.316** | 0.374** | 0.347** | 0.218** | 0.196** | 0.298** | 0.351** | 0.237** |
| 17. WIF-N | −0.063 | −0.018 | 0.407** | 0.462** | 0.426** | 0.342** | 0.415** | 0.395** | 0.278** | 0.190** | 0.271** | 0.362** | 0.236** |
| 18. ITBS9 | −0.066 | 0.127* | 0.610** | 0.390** | 0.400** | 0.284** | 0.381** | 0.325** | 0.195** | 0.459** | 0.446** | 0.391** | 0.330** |
| 19. ITBS10 | −0.067 | 0.096 | 0.539** | 0.364** | 0.356** | 0.246** | 0.316** | 0.221** | 0.174** | 0.383** | 0.386** | 0.259** | 0.271** |
| 20. VSP | −0.047 | −0.002 | 0.541** | 0.430** | 0.404** | 0.236** | 0.364** | 0.272** | 0.158** | 0.397** | 0.461** | 0.402** | 0.347** |

| Variables | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|
| 14. WASI Vocabulary | – | | | | | | |
| 15. WDRB LC | 0.603** | – | | | | | |
| 16. WIF-B | 0.249** | 0.330** | – | | | | |
| 17. WIF-N | 0.224** | 0.322** | 0.892** | – | | | |
| 18. ITBS9 | 0.403** | 0.414** | 0.430** | 0.422** | – | | |
| 19. ITBS10 | 0.352** | 0.365** | 0.311** | 0.322** | 0.561** | – | |
| 20. VSP | 0.271** | 0.344** | 0.448** | 0.429** | 0.602** | 0.512** | – |

*Note.* Across correlations, *N* range 267–340. Age = Age in second grade; DD = Double digit; WASI = Wechsler Abbreviated Scale of Intelligence; MR = Matrix Reasoning; WJ-III = Woodcock-Johnson, Third Edition; CF = Concept Formation; WMTB-C = Working Memory Test Battery for Children; LR = Listening recall; CR = Counting recall; WDRB = Woodcock Diagnostic Reading Battery; LC = Listening comprehension; WIF-B/N = Word identification fluency form B/N; VSP = Vanderbilt Story Problems; ITBS9/10 = Iowa Test of Basic Skills 9/10.
*Note.* Across correlations, *N* range 282–340. WASI = Wechsler Abbreviated Scale of Intelligence; WDRB = Woodcock Diagnostic Reading Battery; LC = Listening comprehension; WIF-B/N = Word identification fluency form B/N; VSP = Vanderbilt Story Problems; ITBS9/10 = Iowa Test of Basic Skills level 9/10.
** $p < .01$
* $p < .05$

For ***girls***, we began by specifying a CFA model identical to the model specified for the full sample (see Fig. 1 and Table C.1). Similar to the full group analysis, this model (Model G1) provided an excellent fit to the data and thus provided the base model for SEM (see Table 4 and Fig. 2a). We initially specified a model that included all variables, which provided a good fit to the data (Model G2; see Tables 4 and 5 for model fit statistics and parameter estimates, respectively). Similar to the full sample, we specified new models after removing the weakest parameters (see Table D.1), which included word reading fluency ($\beta = 0.022$, $p = .816$ in Model G2), concept formation ($\beta = 0.049$, $p = .525$ in Model G3), and working memory ($\beta = 0.200$, $p = .140$ in Model G4). Differences between Models G2 and G3, Models G3 and G4, and Models G4 and G5 were nonsignficant (see Table 4). This final model (Model G5; see Fig. 2b) continued to provide adequate fit to the data ($\chi^2 = 182.665$ (120), $p < .001$, CFI = 0.956, TLI = 0.940, RMSEA = 0.055, 90% confidence interval [0.038, 0.070], p-close = 0.300; SRMR = 0.052). Girls' prior oral language ($\beta = 0.308$, $p < .001$), word-problem solving ($\beta = 0.337$, $p < .001$), computation ($\beta = 0.421$, $p < .001$), and matrix reasoning skills ($\beta = 0.228$, $p < .001$) were all significantly related to word-problem solving in fourth grade; age was not ($\beta = -0.090$, $p = .139$). This model accounted for 85.1% of the variance in fourth-grade word-problem solving performance.

For ***boys***, the CFA model (Model B1) provided an excellent fit (see Fig. 1 and Tables 4 and C.1). Thus, we proceeded to test SEMs using these previously specified latent variables. In the initial SEM, we included all variables and allowed them to correlate (see Fig. 2a). This model (Model B2) provided an excellent fit to the data (see Table 4). As done previously, we removed nonsignificant variables from subsequent models (see Table D.1). These included working memory ($\beta = -0.099$, $p = .555$ in Model B2), concept formation ($\beta = 0.075$, $p = .316$ in Model B3), and word reading fluency ($\beta = 0.093$, $p = .300$ in Model B4). Nonsignificant chi-square differences resulted for all comparisons: Model B2 vs. B3, Model B3 vs. B4, and Model B4 vs. B5 (see Table 4). Thus, we retained the most parsimonious model (Model B5), which provided excellent fit ($\chi^2 = 146.870$ (120), $p = .048$, CFI = 0.979, TLI = 0.971, RMSEA = 0.041, 90% confidence interval [0.004, 0.062], p-close = 0.745; SRMR = 0.049). In this final model, boys' age ($\beta = -0.158$, $p < .05$) and prior oral language ($\beta = 0.424$, $p < .001$), word-problem solving ($\beta = 0.230$, $p < .05$), computation skill ($\beta = 0.381$, $p < .001$), and matrix reasoning ($\beta = 0.301$, $p < .001$) were all significantly associated with fourth-grade word-problem solving (see Table 5 and Fig. 2b). Similar to previous models, these variables accounted for 85.3% variance in fourth-grade word-problem solving.

We also ran ***multi-group models*** to examine these same relations while permitting comparison between boys and girls (see Appendix E for a description of invariance testing for the CFA model). Based on previous modeling, we specified a model that included all constructs as correlated skills, and included only age, word-problem solving, computation skill, matrix reasoning, and oral language as being related to fourth grade word-problem solving. Prior to testing the equivalence of the parameter estimates across the two groups, we first needed to establish equivalence of the means for the observed variables that were not included in the CFA model (age, story problems, matrix reasoning, and concept formation). The unconstrained means model provided a good fit to the data (Model MGc1; see Table 4). A comparison of this model to a model in which the means were constrained to equality (Model MGc2) did not result in a

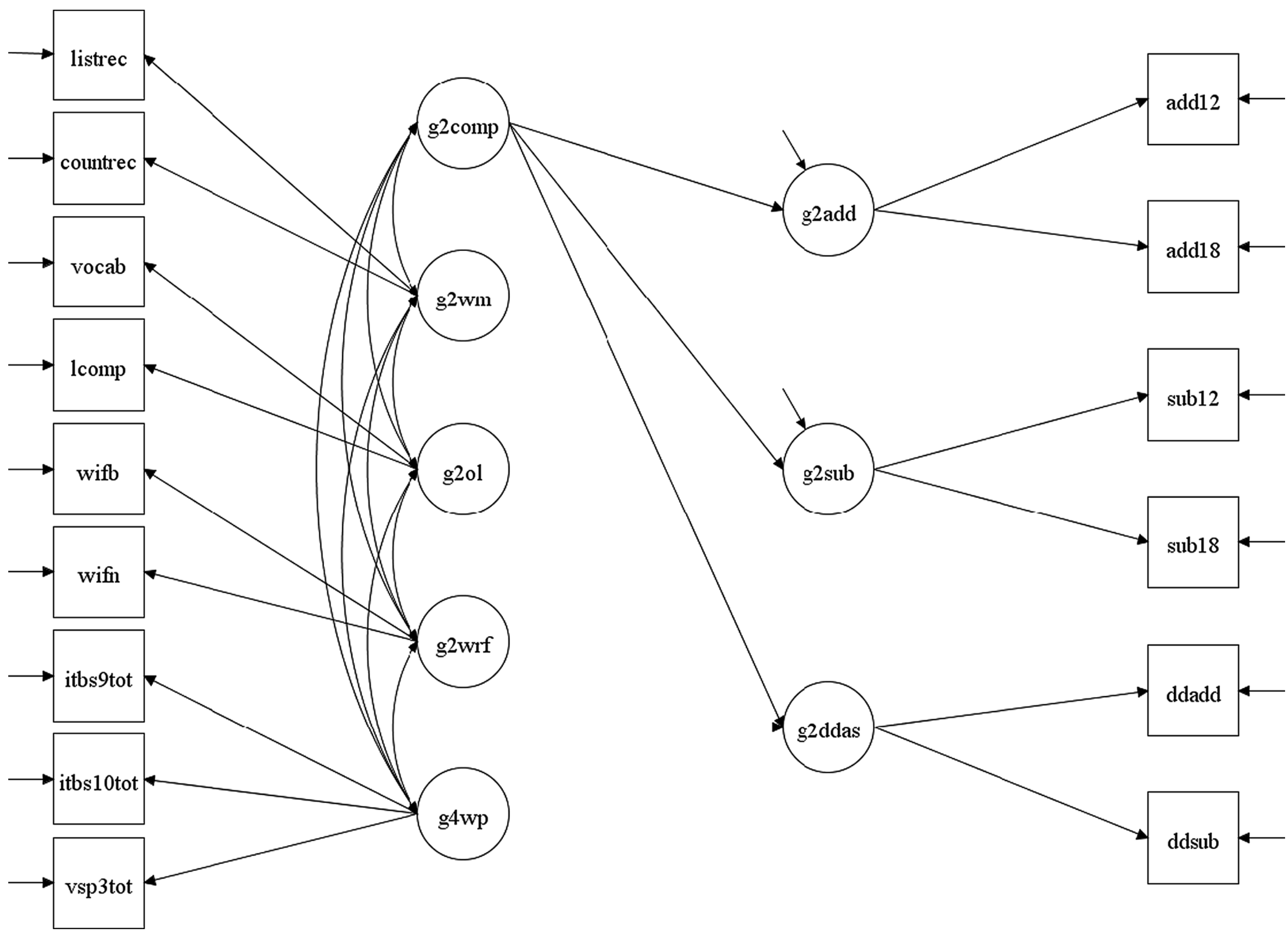


**Fig. 1.** Confirmatory factor analysis model for single- and multi-group modeling. *Note.* G2/4 = Grade 2/4; comp = Computation skill; wm = Working memory; ol = Oral language; wrf = Word reading fluency; wp = Word-problem solving; add = Addition; sub = Subtraction; ddas = Double-digit addition and subtraction; listrec = Listening recall; countrec = Counting recall; vocab = Vocabulary; lcomp = Listening comprehension; wifb/n = Word identification fluency; itbs9/10tot = Iowa Test of Basic Skills 9/10; vsp3tot = Vanderbilt Story Problems.

significant chi-square difference, $\triangle\chi 2 = 8.080$ with 4 *df*, $p = .089$. Therefore, we were able to test for equality for the parameter estimates of the variables across boys and girls.

Constraining the parameter estimates of age (Model MGd1), word-problem solving (Model MGd2), computation skill (Model MGd3), matrix reasoning (Model MGd4), and oral language (Model MGd5) to equality to did not result in significant chi-square differences (see Table 4). Further, the difference between the fully unconstrained (Model MGc2) and the fully constrained model (Model MGd5) was not significant, $\triangle\chi 2 = 4.007$ with 5 *df*, $p = .548$. This suggests that these skills and cognitive processes were not differentially associated for boys versus girls.

## 4. Discussion

In the present study, we examined the longitudinal relations between word-problem solving, computation skill, working memory, nonverbal reasoning, oral language, and word reading fluency in second grade and word-problem solving in fourth grade. As anticipated, foundational word-problem solving and computation skill as well as reasoning ability were significantly associated with later word-problem solving performance in the full sample. Each *SD* increase in second grade word-problem solving, computation skill, and reasoning was associated with a 0.304, 0.386, and 0.276 *SD* increase, respectively, in fourth grade word-problem solving. Moreover, oral language accounted for unique variance in word-problem solving over and above earlier word-problem solving, computation skill, and nonverbal reasoning. For each *SD* increase in oral language, students scored 0.351 *SD* higher on word-problem solving. This indicates that oral language was more strongly related to future word-problem solving than beginning word-problem solving performance.

These results are in line with previous findings showing that foundational skills and oral language ability account for variance in concurrent and longitudinal word-problem solving performance (Fuchs et al., 2006; Fuchs, Zumeta et al., 2010; Tolar et al., 2012; Vilenius-Tuohimaa et al., 2008; Wang et al., 2016; Zheng et al., 2011). Although nonsignificant relations between oral language and word-problem solving have been observed (e.g., Tolar et al., 2012), the present findings are in line with a larger body of work showing that oral language does provide a platform for development word-problem solving (Fuchs et al., 2006, 2015 Fuchs, Fuchs et al., 2008; Wang et al., 2016).

Surprisingly, however, working memory was not significantly related to word-problem solving in any of the models. This stands in contrast to previous studies (e.g., Pina et al., 2014; Wang et al., 2016) and may be due to a few differences between prior investigations and the present study. First, we modeled multiple constructs using a latent variable framework, whereas several studies included only single measures or composite constructs (e.g., Fuchs, Fuchs et al., 2008; Fuchs et al., 2015; Pina et al., 2014). Second, several investigations showing effects for working memory on word-problem solving were concurrent (Fuchs, Fuchs et al., 2008a; Pina et al., 2014) or, if longitudinal designs were used, the gap between assessment periods was shorter than in the present study (Fuchs et al., 2015, 2018; Wang et al., 2016). Third, our working memory latent variable was created using at least one measure that may have tapped language skills (listening recall), and this may be in part why this factor did not account for unique variance over and

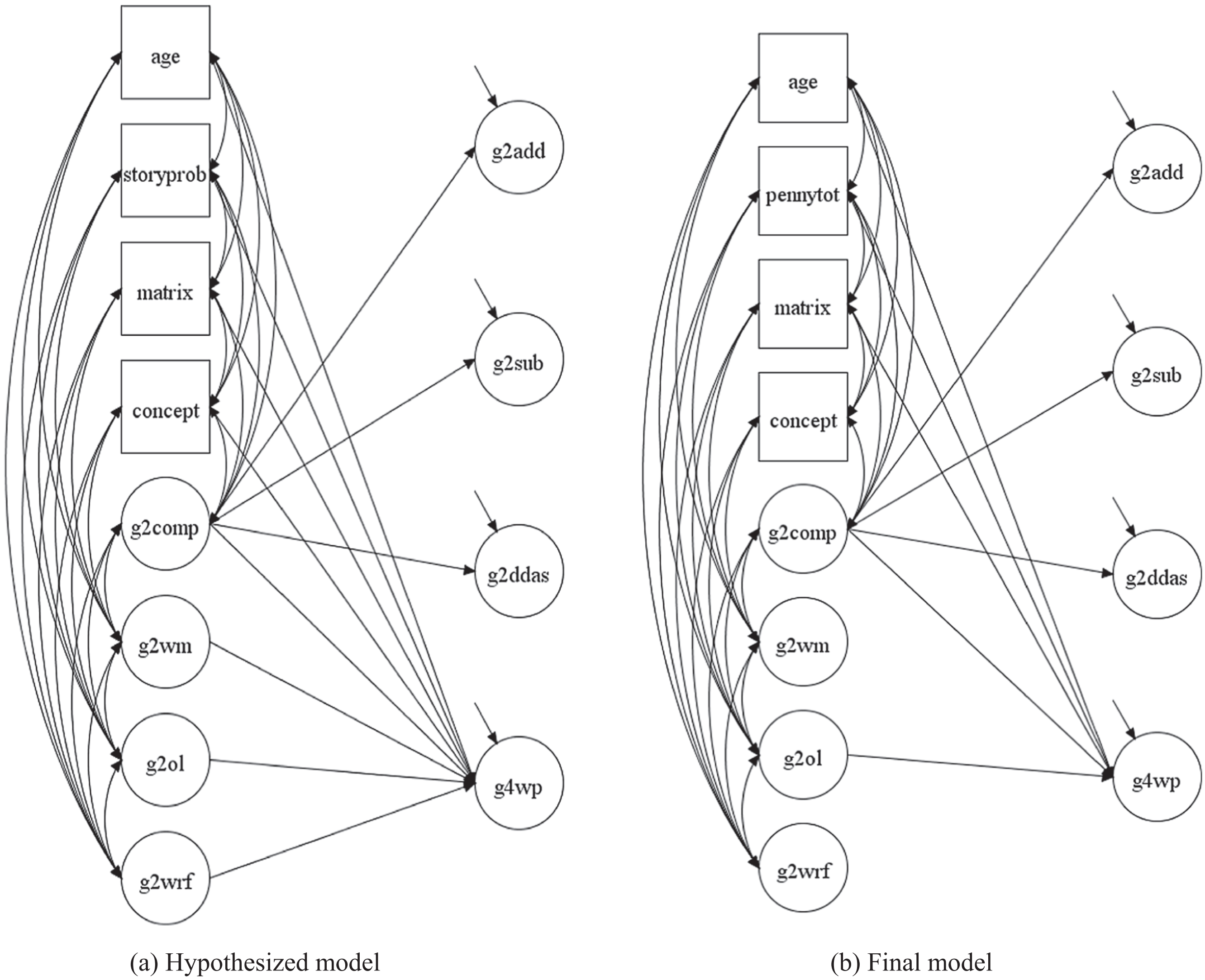


**Fig. 2.** Hypothesized structural equation model (a) and final model (b) for single- and multi-group modeling. *Note.* Observed indicators are omitted. storyprob = Story Problems; matrix = Matrix Reasoning; concept = Concept formation; G2/4 = Grade 2/4; comp = Computation skill; wm = Working memory; ol = Oral language; wrf = Word reading fluency; wp = Word-problem solving; add = Addition; sub = Subtraction; ddas = Double-digit addition and subtraction.

above oral language.

Further, the present sample likely included at least some children who were relatively proficient problem solvers, and the role of working memory in word-problem solving may diminish in proficient problem solvers. Evidence in fact suggests that fluency in mathematical problem solving impacts the extent to which individuals rely on working memory (e.g., Imbo & Vandierendonck, 2007). Thus, given the potentially complex associations between working memory and mathematical problem solving, future investigations should examine potential interactions between mathematical proficiency/automaticity and working memory as well as between language proficiency and working memory and later word-problem solving performance.

When relations were examined for boys and girls separately, we found that these same variables were associated with word-problem solving across both groups. Prior word-problem solving and computation skill appeared to be stronger for girls than for boys (0.337 vs. 0.230 and 0.421 vs. 0.381 for word-problem and computation skill, respectively), while matrix reasoning and oral language seemed to be stronger for boys than for girls (0.301 vs. 0.228 and 0.424 vs. 0.308). However, multigroup modeling indicated these differences were not significant, and although these skills accounted for more variance in word-problem solving for boys than girls, this difference was negligible ($\triangle R^2 = 0.002$). Overall, this pattern suggests that boys and girls do not depend on different sets of cognitive processes and foundational skills to support word-problem solving development.

In terms of theoretical implications, results provide support for the NRC and dual representation models and the problem-model strategy framework of word-problem solving (Hegarty et al., 1995; Kintsch & Greeno, 1985; Mathematics Learning Study Committee, National Research Council, 2001). Finding that early word-problem solving and computation skills is related to later word-problem solving lends support to the assertion that prior knowledge, word-problem solving experience(s), and strategies affect the ability to discern the structure of and successfully solve word-problems (e.g., Kintsch & Greeno, 1985; Mathematics Learning Study Committee, National Research Council, 2001). The association between nonverbal reasoning and word-problem solving also provides support to problem models, in which the ability to solve word problems involves creating a mental problem model and thus requires reasoning ability (Hegarty et al., 1995). The importance of reasoning is well established for mathematics broadly (e.g., mathematical reasoning and logical inference; Geary, 1994; Mathematics Learning Study Committee, National Research Council, 2001) and word-problem solving (Mercer, 1997 cited in Steele, 2002). These relations likely exist because reasoning affects strategy use in mathematics (Casey, Lombardi, Pollock, Fineman, & Pezaris, 2017; Laski et al., 2013) and because the application of strategies impacts word-problem solving (Koedinger & Tabachneck, 1994).

The importance of oral language to word-problem solving is

**Table 4**
Model Comparisons for Single- and Multi-Group Analyses.

| Analysis | Sample | Model | Model Fit Statistics | | | | | | | Model Comparisons | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\chi^2$ | *df* | *p* | CFI | TLI | RMSEA | SRMR | Comparison | $\triangle\chi2$ | $\triangle df$ |
| Single-Group | | | | | | | | | | | | |
| | Full Sample | | | | | | | | | | | |
| | | F1 | 112.781 | 77 | 0.005 | 0.986 | 0.981 | 0.037 | 0.042 | – | – | – |
| | | F2 | 160.190 | 117 | 0.005 | 0.985 | 0.978 | 0.033 | 0.037 | – | – | – |
| | | F3 | 160.610 | 118 | 0.006 | 0.985 | 0.979 | 0.033 | 0.037 | F2 vs. F3 | 0.420 | 1 |
| | | F4 | 161.378 | 119 | 0.006 | 0.985 | 0.979 | 0.032 | 0.037 | F3 vs. F4 | 0.768 | 1 |
| | | F5 | 162.537 | 120 | 0.006 | 0.985 | 0.979 | 0.032 | 0.038 | F4 vs. F5 | 1.159 | 1 |
| | Girls | | | | | | | | | | | |
| | | G1 | 115.388 | 77 | 0.003 | 0.969 | 0.958 | 0.054 | 0.057 | – | – | – |
| | | G2 | 179.815 | 117 | < 0.001 | 0.956 | 0.938 | 0.056 | 0.051 | – | – | – |
| | | G3 | 179.869 | 118 | < 0.001 | 0.957 | 0.939 | 0.055 | 0.051 | G2 vs. G3 | 0.054 | 1 |
| | | G4 | 180.270 | 119 | < 0.001 | 0.957 | 0.940 | 0.054 | 0.051 | G3 vs. G4 | 0.401 | 1 |
| | | G5 | 182.665 | 120 | < 0.001 | 0.956 | 0.940 | 0.055 | 0.052 | G4 vs. G5 | 2.395 | 1 |
| | Boys | | | | | | | | | | | |
| | | B1 | 101.318 | 77 | 0.033 | 0.978 | 0.970 | 0.048 | 0.052 | – | – | – |
| | | B2 | 144.455 | 117 | 0.043 | 0.978 | 0.969 | 0.042 | 0.048 | – | – | – |
| | | B3 | 144.830 | 118 | 0.047 | 0.979 | 0.970 | 0.041 | 0.048 | B2 vs. B3 | 0.375 | 1 |
| | | B4 | 145.831 | 119 | 0.048 | 0.979 | 0.970 | 0.041 | 0.049 | B3 vs. B4 | 1.001 | 1 |
| | | B5 | 146.870 | 120 | 0.048 | 0.979 | 0.971 | 0.041 | 0.049 | B4 vs. B5 | 1.039 | 1 |
| Multi-Group | | | | | | | | | | | | |
| | Girls/Boys | | | | | | | | | | | |
| | | MGa1 | 187.557 | 138 | 0.003 | 0.979 | 0.968 | 0.048 | 0.046 | – | – | – |
| | | MGa2 | 201.999 | 146 | 0.002 | 0.976 | 0.966 | 0.050 | 0.051 | MGa1 vs. MGa2 | 14.442 | 8 |
| | | MGa3 | 214.984 | 154 | 0.001 | 0.974 | 0.965 | 0.051 | 0.052 | MGa2 vs. MGa3 | 12.985 | 8 |
| | | MGb1 | 246.363 | 170 | < 0.001 | 0.968 | 0.960 | 0.054 | 0.060 | – | – | – |
| | | MGb2 | 247.726 | 172 | < 0.001 | 0.968 | 0.961 | 0.053 | 0.061 | MGb1 vs. MGb2 | 1.363 | 2 |
| | | MGb3 | 258.686 | 179 | < 0.001 | 0.966 | 0.961 | 0.054 | 0.064 | MGb3 vs. MGb2 | 10.960 | 7 |
| | | MGc1 | 366.476 | 265 | < 0.001 | 0.962 | 0.953 | 0.050 | 0.057 | – | – | – |
| | | MGc2 | 374.556 | 269 | < 0.001 | 0.961 | 0.952 | 0.050 | 0.059 | MGc1 vs. MGc2 | 8.080 | 4 |
| | | MGd1 | 374.583 | 270 | < 0.001 | 0.961 | 0.952 | 0.050 | 0.059 | – | – | – |
| | | MGd2 | 375.229 | 271 | < 0.001 | 0.961 | 0.953 | 0.050 | 0.059 | MGd1 vs. MGd2 | 0.646 | 1 |
| | | MGd3 | 377.444 | 272 | < 0.001 | 0.961 | 0.952 | 0.050 | 0.061 | MGd2 vs. MdG3 | 2.215 | 1 |
| | | MGd4 | 378.550 | 273 | < 0.001 | 0.961 | 0.952 | 0.050 | 0.061 | MGd3 vs. MGd4 | 1.106 | 1 |
| | | MGd5 | 378.563 | 274 | < 0.001 | 0.961 | 0.953 | 0.050 | 0.061 | MGd4 vs. MGd5 | 0.013 | 1 |

*Note.* F = Full sample; G = Girls; B = Boys; Models F/G/B1 are confirmatory factor analysis (CFA) models; F/B/G2-5 are structural equation models (SEM); MG = Multigroup; a = Testing first-order CFA factors; b = Testing second-order CFA factors; c = Testing SEM means for observed variables; d = Testing SEM path constraints; $\chi2$ = Chi-square; *df* = Degrees of freedom; CFI = Comparative fit index; TLI = Tucker-Lewis index; RMSEA = Root mean square error of approximation; SRMR = Standardized root mean square residual.

**Table 5**
Parameter Estimates for Final Models of Fourth-Grade Word-Problem Solving.

| Sample | Model | Variables | Estimate | *SE* |
|---|---|---|---|---|
| Full Sample | | | | |
| | F5 | | | |
| | | Age | −0.107* | 0.042 |
| | | Story Problems | 0.304** | 0.061 |
| | | Computation Skill | 0.386** | 0.056 |
| | | Matrix Reasoning | 0.276** | 0.045 |
| | | Oral Language | 0.351** | 0.056 |
| Girls | | | | |
| | G5 | | | |
| | | Age | −0.090 | 0.061 |
| | | Story Problems | 0.337** | 0.081 |
| | | Computation Skill | 0.421** | 0.082 |
| | | Matrix Reasoning | 0.228** | 0.065 |
| | | Oral Language | 0.308** | 0.080 |
| Boys | | | | |
| | B5 | | | |
| | | Age | −0.158* | 0.063 |
| | | Story Problems | 0.230* | 0.102 |
| | | Computation Skill | 0.381** | 0.091 |
| | | Matrix Reasoning | 0.301** | 0.067 |
| | | Oral Language | 0.424** | 0.089 |

*Note.* Estimates are standardized. F = Full sample ($N$ = 340); G = Girls ($N$ = 174); B = Boys ($N$ = 135); *SE* = Standard error.
* $p < .05$.
** $p < .001$.

additionally emphasized in both models of word-problem solving (Hegarty et al., 1995; Kintsch & Greeno, 1985). These models assert that when presented with a word problem, individuals construct a text base that facilitates an understanding of the text, thus leading to development of the problem model. That is, the ability to understand the text is supported by an understanding of individual words (e.g., Hoover & Gough, 1990) in addition to the language structure. Further, the strength of the relation between oral language and word-problem solving, in the context of a rich model with many competing foundational skills and cognitive abilities, lends additional credence to Kintsch and Greeno's (1985) hypothesis that word-problem solving is at least, in part, a form of text comprehension.

In terms of practical implications, the present findings are in line with previous investigations showing that children who have difficulty with word-problem solving, computation, nonverbal reasoning, and oral language in second grade likely endure difficulty in solving word problems over time. Given this, interventions that aim to build these foundational skills in addition to building reasoning and language processes in the context of word-problem solving instruction are needed to remediate word-problem solving difficulties early in school (Fuchs & Fuchs, 2002; Fuchs, Seethaler et al., 2008; Kong & Orosco, 2016; Träf & Samuelsson, 2013).

Although the present study suggests that researchers may not need to be sensitive to sex differences when selecting measures for identifying students for such services, further research on this topic is needed because the absence of significant differences between boys and girls may be partly due to insufficient power. With larger samples, researchers may detect some of the differences suggested in the observed magnitudes in relations for boys versus girls (Δβ were 0.107 and 0.040 in foundational word-problem solving and computation, respectively, favoring girls; Δβ were 0.073 and 0.116 for matrix reasoning and oral language, respectively, favoring boys).

Before closing, we note four study limitations. First, we remind readers that the present set of analyses, although prospectively longitudinal, are correlational. Therefore, causal inferences are not possible based on present findings. Second, our sample includes the full distribution children. As a result, we cannot draw conclusions specifically for children with learning disabilities or other disorders. Third, although we included measures of various cognitive processes, we did not include some potentially associative constructs (e.g., attention and reading comprehension; Björn et al., 2016; Fuchs et al., 2006). It is possible that inclusion of these variables within a single model may change the observed pattern of results. Fourth, findings apply only to students in the early elementary grades. In the later grades, the cognitive demands of more complex word-problem solving may alter the nature of associations. Nonetheless, the present investigation adds to our understanding of the developmental relations between children's early cognitive skills and later word-problem solving.

### 4.1. Conclusions

Findings suggest that early language-related skills, such as vocabulary and language comprehension, are significantly associated with children's later word-problem solving outcomes over and above prior word-problem solving, computational abilities, and nonverbal reasoning. Because the observed role of oral language in word-problem solving aligns with several theoretical frameworks of word-problem solving, such as the dual representation model, it is necessary to consider language skills in conjunction with children's reasoning and mathematical abilities when designing intervention programs for remediating poor word-problem solving outcomes. Results also indicate similar associations across boys and girls, although additional research is warranted. In sum, this investigation adds important information about individual differences in word-problem solving development and suggests that language skills should be considered within word-problem solving interventions.

## Declaration of Competing Interest

None.

## Acknowledgement

This research was supported by Award Numbers 3 R01 HD044073-14S1, 2 P20 HD075443, R01 HD053714, 5 R01 HD044073, and Core Grant HD15052 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development to Vanderbilt University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health.

## Appendix A

See Table A.1

**Table A.1**
Descriptive Statistics for Girls and Boys.

| Measure | Girls | | | | | | | Boys | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *N* | Min. | Max. | Mean | *SD* | Skew | Kurtosis | *N* | Min. | Max. | Mean | *SD* | Skew | Kurtosis |
| Age | 174 | 6.10 | 8.81 | 7.585 | 0.425 | 0.007 | 1.107 | 135 | 6.69 | 9.02 | 7.681 | 0.441 | 0.431 | −0.067 |
| Story Problems[a] | 174 | 0 | 14 | 6.138 | 3.309 | 0.280 | −0.599 | 134 | 1 | 14 | 7.007 | 3.534 | 0.337 | −0.841 |
| Addition 0–12[a] | 174 | 2 | 22 | 8.316 | 4.565 | 0.820 | 0.299 | 133 | 1 | 22 | 8.722 | 4.887 | 0.615 | −0.201 |
| Addition 0–18[a] | 174 | 0 | 16 | 6.017 | 3.547 | 0.548 | −0.122 | 133 | 0 | 16 | 6.226 | 3.513 | 0.494 | 0.060 |
| DD Addition[a] | 173 | 0 | 12 | 3.254 | 3.024 | 0.726 | −0.356 | 132 | 0 | 12 | 3.212 | 3.111 | 0.908 | 0.062 |
| Subtraction 0–12[a] | 174 | 0 | 12 | 4.052 | 2.499 | 0.633 | 0.002 | 133 | 0 | 12 | 4.744 | 2.811 | 0.586 | 0.165 |
| Subtraction 0–18[a] | 174 | 0 | 10 | 2.885 | 2.229 | 0.805 | 0.101 | 133 | 0 | 11 | 3.368 | 2.610 | 0.975 | 0.874 |
| DD Subtraction[a] | 173 | 0 | 9 | 2.358 | 2.597 | 0.992 | 0.166 | 132 | 0 | 9 | 2.053 | 2.662 | 1.093 | 0.079 |
| WASI MR | 174 | 2 | 26 | 11.270 | 5.686 | 0.727 | −0.582 | 135 | 3 | 28 | 12.000 | 5.869 | 0.591 | −0.518 |
| WJ-CF | 174 | 3 | 27 | 11.253 | 5.492 | 0.333 | −0.597 | 135 | 1 | 27 | 11.489 | 6.028 | 0.767 | 0.226 |
| WMTB-C LR | 174 | 0 | 16 | 7.121 | 3.838 | −0.010 | −0.538 | 135 | 0 | 16 | 7.215 | 3.724 | −0.244 | −0.236 |
| WMTB-C CR | 174 | 4 | 24 | 13.977 | 4.394 | 0.053 | −0.390 | 135 | 2 | 26 | 14.341 | 4.678 | 0.175 | 0.031 |
| WASI Vocabulary | 174 | 7 | 37 | 22.149 | 5.938 | 0.126 | 0.188 | 135 | 5 | 37 | 21.919 | 7.089 | −0.037 | −0.368 |
| WDRB LC | 174 | 5 | 29 | 16.172 | 4.495 | −0.382 | 0.097 | 135 | 5 | 26 | 17.015 | 4.985 | −0.441 | −0.236 |
| WIF-B | 174 | 0 | 111 | 41.667 | 24.485 | 0.387 | −0.594 | 135 | 2 | 111 | 41.935 | 24.836 | 0.385 | −0.505 |
| WIF-N | 173 | 1 | 120 | 56.543 | 22.107 | −0.171 | 0.046 | 135 | 7 | 105 | 55.733 | 22.687 | −0.070 | −0.612 |
| ITBS9[b] | 152 | 1 | 17 | 9.711 | 3.816 | −0.121 | −0.730 | 115 | 2 | 17 | 10.670 | 3.573 | −0.270 | −0.496 |
| ITBS10[b] | 152 | 1 | 11 | 5.257 | 2.073 | 0.315 | −0.363 | 115 | 1 | 11 | 5.678 | 2.308 | 0.332 | −0.588 |
| VSP[b] | 152 | 0 | 29 | 11.066 | 6.756 | 0.541 | −0.199 | 115 | 0 | 29 | 11.043 | 6.719 | 0.684 | 0.246 |

*Note.* Raw scores reported; *SD* = Standard deviation; Min. = Minimum; Max. = Maximum; Age = Age in second grade; DD = Double digit; WASI = Wechsler Abbreviated Scale of Intelligence; MR = Matrix Reasoning; WJ = III = Woodcock-Johnson; CF = Concept Formation; WMTB-C = Working Memory Test Battery for Children; L/CR = Listening/counting recall; WDRB = Woodcock Diagnostic Reading Battery; LC = Listening comprehension; WIF-B/N = Word identification fluency form B/N; VSP = Vanderbilt Story Problems; ITBS9/10 = Iowa Test of Basic Skills level 9/10.
[a] Timed.
[b] Outcome.

## Appendix B

See Table B.1

**Table B.1**
Correlations Across Girls and Boys.

| Group | Measure | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Girls | | | | | | | | | | | | | |
| | 1. Age | – | | | | | | | | | | | |
| | 2. Story Problems | 0.070 | – | | | | | | | | | | |
| | 3. Addition 0–12 | 0.178* | 0.268** | – | | | | | | | | | |
| | 4. Addition 0–18 | 0.086 | 0.264** | 0.754** | – | | | | | | | | |
| | 5. DD Addition | 0.074 | 0.240** | 0.522** | 0.492** | – | | | | | | | |
| | 6. Subtraction 0–12 | 0.149 | 0.363** | 0.502** | 0.549** | 0.290** | – | | | | | | |
| | 7. Subtraction 0–18 | 0.193* | 0.330** | 0.427** | 0.399** | 0.264** | 0.740** | – | | | | | |
| | 8. DD Subtraction | −0.087 | 0.262** | 0.465** | 0.415** | 0.730** | 0.153* | 0.206** | – | | | | |
| | 9. WASI MR | 0.005 | 0.361** | 0.206** | 0.311** | 0.217** | 0.224** | 0.183* | 0.243** | – | | | |
| | 10. WJ-III CF | 0.012 | 0.463** | 0.239** | 0.241** | 0.242** | 0.222** | 0.227** | 0.228** | 0.358** | – | | |
| | 11. WMTB-C LR | 0.059 | 0.425** | 0.157* | 0.171* | 0.190* | 0.158* | 0.171* | 0.205** | 0.326** | 0.392** | – | |
| | 12. WMTB-C CR | 0.069 | 0.249** | 0.220** | 0.221** | 0.283** | 0.123 | 0.138 | 0.317** | 0.273** | 0.300** | 0.408** | – |
| | 13. WASI Vocabulary | 0.011 | 0.353** | 0.072 | 0.132 | 0.004 | −0.027 | −0.038 | 0.021 | 0.191* | 0.385** | 0.366** | 0.221** |
| | 14. WDRB LC | 0.071 | 0.397** | 0.135 | 0.163* | 0.086 | 0.173* | 0.169* | 0.125 | 0.189* | 0.443** | 0.474** | 0.111 |
| | 15. WIF-B | 0.079 | 0.427** | 0.340** | 0.338** | 0.265** | 0.410** | 0.376** | 0.199** | 0.259** | 0.254** | 0.398** | 0.274** |
| | 16. WIF-N | −0.003 | 0.406** | 0.404** | 0.398** | 0.314** | 0.475** | 0.441** | 0.286** | 0.258** | 0.251** | 0.406** | 0.277** |
| | 17. ITBS9 | 0.009 | 0.583** | 0.405** | 0.400** | 0.317** | 0.354** | 0.308** | 0.267** | 0.454** | 0.418** | 0.447** | 0.253** |
| | 18. ITBS10 | −0.040 | 0.489** | 0.372** | 0.331** | 0.292** | 0.301** | 0.246** | 0.268** | 0.346** | 0.343** | 0.260** | 0.232** |
| | 19. VSP | 0.038 | 0.536** | 0.311** | 0.298** | 0.197* | 0.278** | 0.198* | 0.125 | 0.368** | 0.437** | 0.443** | 0.292** |
| Group | Measure | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Boys | | | | | | | | | | | | | |
| | 1. Age | – | | | | | | | | | | | |
| | 2. Story Problems | −0.077 | – | | | | | | | | | | |
| | 3. Addition 0–12 | −0.066 | 0.478** | – | | | | | | | | | |
| | 4. Addition 0–18 | −0.085 | 0.516** | 0.801** | – | | | | | | | | |
| | 5. DD Addition | −0.110 | 0.319** | 0.518** | 0.494** | – | | | | | | | |
| | 6. Subtraction 0–12 | −0.001 | 0.398** | 0.649** | 0.543** | 0.399** | – | | | | | | |
| | 7. Subtraction 0–18 | −0.039 | 0.347** | 0.653** | 0.576** | 0.409** | 0.754** | – | | | | | |
| | 8. DD Subtraction | −0.011 | 0.223* | 0.478** | 0.463** | 0.704** | 0.348** | 0.425** | – | | | | |
| | 9. WASI MR | −0.023 | 0.199* | 0.119 | 0.167 | −0.057 | 0.165 | 0.129 | 0.011 | – | | | |
| | 10. WJ-III CF | −0.011 | 0.494** | 0.254** | 0.301** | 0.096 | 0.254** | 0.229** | 0.045 | 0.279** | – | | |
| | 11. WMTB-C LR | −0.086 | 0.417** | 0.257** | 0.172* | 0.180* | 0.165 | 0.126 | −0.020 | 0.236** | 0.234** | – | |
| | 12. WMTB-C CR | −0.031 | 0.341** | 0.354** | 0.266** | 0.290** | 0.311** | 0.314** | 0.161 | 0.313** | 0.254** | 0.405** | – |
| | 13. WASI Vocabulary | 0.061 | 0.440** | 0.071 | 0.022 | 0.040 | −0.024 | −0.026 | −0.073 | 0.161 | 0.365** | 0.243** | 0.231** |
| | 14. WDRB LC | 0.002 | 0.410** | 0.105 | 0.088 | 0.018 | 0.046 | 0.070 | −0.133 | 0.177* | 0.372** | 0.338** | 0.329** |
| | 15. WIF-B | −0.176* | 0.391** | 0.445** | 0.402** | 0.382** | 0.335** | 0.305** | 0.232** | 0.057 | 0.346** | 0.308** | 0.189* |
| | 16. WIF-N | −0.116 | 0.383** | 0.515** | 0.434** | 0.403** | 0.356** | 0.350** | 0.280** | 0.057 | 0.291** | 0.341** | 0.200* |
| | 17. ITBS9 | −0.189* | 0.640** | 0.397** | 0.418** | 0.287** | 0.416** | 0.352** | 0.166 | 0.444** | 0.509** | 0.309** | 0.412** |
| | 18. ITBS10 | −0.078 | 0.584** | 0.366** | 0.399** | 0.217* | 0.330** | 0.216* | 0.129 | 0.430** | 0.436** | 0.269** | 0.321** |
| | 19. VSP | −0.128 | 0.551** | 0.562** | 0.552** | 0.291** | 0.486** | 0.384** | 0.227* | 0.416** | 0.504** | 0.359** | 0.411** |

| Group | Measure | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|
| Girls | | | | | | | | |
| | 13. WASI Vocabulary | – | | | | | | |
| | 14. WDRB LC | 0.578** | – | | | | | |
| | 15. WIF-B | 0.242** | 0.291** | – | | | | |
| | 16. WIF-N | 0.232** | 0.333** | 0.883** | – | | | |
| | 17. ITBS9 | 0.376** | 0.376** | 0.394** | 0.403** | – | | |
| | 18. ITBS10 | 0.312** | 0.350** | 0.272** | 0.321** | 0.480** | – | |
| | 19. VSP | 0.258** | 0.360** | 0.450** | 0.447** | 0.589** | 0.440** | – |
| Boys | | | | | | | | |
| | 13. WASI Vocabulary | – | | | | | | |
| | 14. WDRB LC | 0.652** | – | | | | | |
| | 15. WIF-B | 0.217* | 0.317** | – | | | | |
| | 16. WIF-N | 0.177* | 0.255** | 0.900** | – | | | |
| | 17. ITBS9 | 0.478** | 0.452** | 0.492** | 0.482** | – | | |
| | 18. ITBS10 | 0.413** | 0.409** | 0.366** | 0.346** | 0.665** | – | |
| | 19. VSP | 0.293** | 0.354** | 0.437** | 0.408** | 0.633** | 0.606** | – |

*Note.* Across correlations, *N* ranges 151–174 and 112–135 for girls and boys, respectively. Age = Age in second grade; DD = Double digit; WASI = Wechsler Abbreviated Scale of Intelligence; MR = Matrix Reasoning; WJ = III = Woodcock-Johnson, Third Edition; CF = Concept Formation; WMTB-C = Working Memory Test Battery for Children; LR = Listening recall; CR = Counting recall; Vocab. = Vocabulary; WDRB = Woodcock Diagnostic Reading Battery; LC = Listening comprehension; WIF-B/N = Word identification fluency form B/N; VSP = Vanderbilt Story Problems; ITBS9/10 = Iowa Test of Basic Skills level 9/10.

*Note.* Across correlations, *N* ranges 151–174 and 112–135 for girls and boys, respectively. Age = Age in second grade; DD = Double digit; WASI = Wechsler Abbreviated Scale of Intelligence; MR = Matrix Reasoning; WJ = III = Woodcock-Johnson, Third Edition; CF = Concept Formation; WMTB-C = Working Memory Test Battery for Children; LR = Listening recall; CR = Counting recall; WDRB = Woodcock Diagnostic Reading Battery; LC = Listening comprehension; WIF-B/N = Word identification fluency form B/N; VSP = Vanderbilt Story Problems; ITBS9/10 = Iowa Test of Basic Skills level 9/10.

*Note.* Across correlations, *N* ranges 151–174 and 115–135 for girls and boys, respectively WASI = Wechsler Abbreviated Scale of Intelligence; WDRB = Woodcock Diagnostic Reading Battery; LC = Listening comprehension; WIF-B/N = Word identification fluency form B/N; VSP = Vanderbilt Story Problems; ITBS9/10 = Iowa Test of Basic Skills level 9/10.

** $p < .01$.

* $p < .05$.

## Appendix C

See Table C.1

**Table C.1**
Factor Loadings for Confirmatory Factor Analysis Models for the Full Sample (N = 340), Girls Only (N = 174), and Boys Only (N = 135).

| Latent Variable | Indicators | Factor Loading | | |
|---|---|---|---|---|
| | | Full Sample | Girls | Boys |
| G2 Addition | | | | |
| | Addition 0–12 | 0.906 | 0.875 | 0.945 |
| | Addition 0–18 | 0.839 | 0.862 | 0.849 |
| G2 Subtraction | | | | |
| | Subtraction 0–12 | 0.909 | 0.945 | 0.871 |
| | Subtraction 0–18 | 0.825 | 0.783 | 0.868 |
| G2 DD Addition/ Subtraction | | | | |
| | DD Addition | 0.920 | 0.946 | 0.887 |
| | DD Subtraction | 0.781 | 0.772 | 0.794 |
| G2 Computation Skill | | | | |
| | G2 Addition | 0.933 | 0.911 | 0.949 |
| | G2 Subtraction | 0.738 | 0.695 | 0.827 |
| | G2 DD Addition/ Subtraction | 0.654 | 0.633 | 0.673 |
| G2 Working Memory | | | | |
| | WMTB-C LR | 0.745 | 0.868 | 0.595 |
| | WMTB-C CR | 0.525 | 0.470 | 0.681 |
| G2 Word Reading Fluency | | | | |
| | WIF-B | 0.915 | 0.894 | 0.929 |
| | WIF-N | 0.974 | 0.988 | 0.969 |
| G2 Oral Language | | | | |
| | WASI Vocabulary | 0.724 | 0.684 | 0.788 |
| | WDRB LC | 0.833 | 0.846 | 0.828 |
| G4 Word Problem Solving | | | | |
| | ITBS9 | 0.810 | 0.806 | 0.837 |
| | ITBS10 | 0.686 | 0.625 | 0.760 |
| | VSP | 0.755 | 0.732 | 0.782 |

*Note.* Factor loadings are standardized estimates. G2/4 = Grade 2/4; DD = Double-digit; WMTB-C = Working Memory Test Battery for Children; WIF-B/N = Word identification fluency Form B/N; WASI = Wechsler Abbreviated Scale of Intelligence; WDRB = Woodcock Diagnostic Reading Battery; LC = Listening comprehension; ITBS9/10 = Iowa Test of Basic Skills level 9/10; VSP = Vanderbilt Story Problems. All loadings are significant at $p < .001$.

## Appendix D

See Table D.1

**Table D.1**
Parameter Estimates Across All Models Predicting Fourth-Grade Word-Problem Solving.

| Sample | Model | Variable | Estimate | *SE* |
|---|---|---|---|---|
| Full Sample | | | | |
| | F2 | | | |
| | | Age | −0.095* | 0.043 |
| | | Story Problems | 0.300*** | 0.060 |
| | | Computation Skill | 0.325*** | 0.072 |
| | | Matrix Reasoning | 0.259*** | 0.049 |
| | | Concept Formation | 0.067 | 0.052 |
| | | Working Memory | 0.062 | 0.096 |
| | | Oral Language | 0.262** | 0.082 |
| | | Word Reading Fluency | 0.046 | 0.064 |
| | F3 | | | |
| | | Age | −0.095* | 0.043 |
| | | Story Problems | 0.307*** | 0.059 |
| | | Computation Skill | 0.334*** | 0.072 |
| | | Matrix Reasoning | 0.272*** | 0.045 |
| | | Concept Formation | 0.066 | 0.052 |
| | | Oral Language | 0.292*** | 0.069 |
| | | Word Reading Fluency | 0.055 | 0.062 |
| | F4 | | | |
| | | Age | −0.103* | 0.042 |
| | | Story Problems | 0.300*** | 0.060 |
| | | Computation Skill | 0.375*** | 0.057 |

(continued on next page)

**Table D.1** (*continued*)

| Sample | Model | Variable | Estimate | *SE* |
|---|---|---|---|---|
| | | Matrix Reasoning | 0.271*** | 0.045 |
| | | Concept Formation | 0.056 | 0.052 |
| | | Oral Language | 0.321*** | 0.062 |
| Girls | | | | |
| | G2 | | | |
| | | Age | −0.078 | 0.062 |
| | | Story Problems | 0.328*** | 0.079 |
| | | Computation Skill | 0.356** | 0.107 |
| | | Matrix Reasoning | 0.192** | 0.069 |
| | | Concept Formation | 0.055 | 0.081 |
| | | Working Memory | 0.181 | 0.136 |
| | | Oral Language | 0.156 | 0.117 |
| | | Word Reading Fluency | 0.022 | 0.095 |
| | G3 | | | |
| | | Age | −0.082 | 0.060 |
| | | Story Problems | 0.326*** | 0.079 |
| | | Computation Skill | 0.372*** | 0.085 |
| | | Matrix Reasoning | 0.190** | 0.068 |
| | | Working Memory | 0.193 | 0.131 |
| | | Concept Formation | 0.049 | 0.077 |
| | | Oral Language | 0.161 | 0.116 |
| | G4 | | | |
| | | Age | −0.086 | 0.060 |
| | | Story Problems | 0.326*** | 0.079 |
| | | Computation Skill | 0.380*** | 0.085 |
| | | Working Memory | 0.200 | 0.136 |
| | | Matrix Reasoning | 0.195** | 0.069 |
| | | Oral Language | 0.186 | 0.112 |
| Boys | | | | |
| | B2 | | | |
| | | Age | −0.146* | 0.066 |
| | | Story Problems | 0.251* | 0.105 |
| | | Computation Skill | 0.341* | 0.133 |
| | | Matrix Reasoning | 0.328*** | 0.081 |
| | | Concept Formation | 0.064 | 0.078 |
| | | Working Memory | −0.099 | 0.167 |
| | | Oral Language | 0.397** | 0.134 |
| | | Word Reading Fluency | 0.094 | 0.091 |
| | B3 | | | |
| | | Age | −0.140* | 0.064 |
| | | Story Problems | 0.239* | 0.100 |
| | | Computation Skill | 0.300** | 0.107 |
| | | Matrix Reasoning | 0.301*** | 0.066 |
| | | Concept Formation | 0.075 | 0.075 |
| | | Oral Language | 0.348*** | 0.100 |
| | | Word Reading Fluency | 0.095 | 0.089 |
| | B4 | | | |
| | | Age | −0.143* | 0.064 |
| | | Story Problems | 0.254* | 0.101 |
| | | Computation Skill | 0.314** | 0.108 |
| | | Matrix Reasoning | 0.308*** | 0.066 |
| | | Oral Language | 0.379*** | 0.097 |
| | | Word Reading Fluency | 0.093 | 0.090 |

*Note.* Estimates are standardized. *SE* = Standard error.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

## Appendix E

Invariance Testing for Multi-Group Confirmatory Factor Analysis (CFA) Modeling.

Although we previously established that the latent constructs were adequately represented by their observed indicators, the multi-group CFA models allowed for a test of whether the means and factor loadings were similar or dissimilar across the two groups. To test this, we specified three different models: a configural model, a metric model, and a scalar model. In the configural model, the observed variable means and the factor loadings are allowed to vary; essentially, we are testing whether the factor structure is equivalent across boys and girls. In the metric model, we imposed constraints on the factor loadings; however, the means are allowed to vary across the two groups. Finally, in the scalar model, we constrained the means and factor loadings to equality across groups. Invariance across both means and factor loadings is necessary to meaningfully compare the two groups. Given that we included a higher-order factor within our CFA models, we first tested invariance for the first-order factors and then followed-up with invariance testing for the higher-order factor separately.

The configural model (Model MGa1) provided an excellent fit to the data (see Table 4), indicating the factor structure was appropriately represented for the first-order factors for boys and girls. Following this, we tested for metric invariance. Because the configural model and the metric model (Model MGa2) are nested, we used chi-square difference testing to determine whether adding equality constraints led to a significant worsening of model fit. When compared to the configural model, imposing equality constraints on the factor loadings did not result in a significantly worse fitting model, $\triangle\chi2 = 14.442$ with 8 *df*, $p = .071$, suggesting the factor loadings were similar across the two groups. Thus, the more parsimonious metric model was retained. Comparing the metric model to the scalar model (Model MGa3), while imposing additional equality constraints on the means across the two groups, did not result in significantly worse fit, $\triangle\chi2 = 12.985$ with 8 *df*, $p = .112$. So the scalar model was the most parsimonious and preferred model. Thus, we established that the mean structure and factor loadings were equivalent for the first-order factors across boys and girls.

Because we were able to establish invariance for the first-order factors, we tested for metric invariance for the second-order factor (i.e., whether the loadings for the subtraction and addition factors were similar across the two groups). For the multi-group models, we fixed the unstandardized path from the computation factor to one of the first order factors to 1 given that standardization is inappropriate in multi-group analyses (see Kline, 2011). First, we specified a configural model for the factor structure of the second-order factor only (MGb1). Following this, we constrained the factor loadings to equality across boys and girls (MGb2). A comparison of this model to the configural model resulted in a nonsignificant chi-square difference, $\triangle\chi2 = 1.363$ with 2 *df*, $p = .506$. Scalar invariance was tested by constraining the latent factor means to equality (i.e., fixed to zero across groups; Model MGb3). This did not lead to a significant degradation in model fit compared to the metric model, $\triangle\chi2 = 10.960$ with 7 *df*, $p = .140$. Because we established invariance for the full CFA model, we proceeded by testing equivalence of the estimates for the best-fitting SEMs for boys and girls (across Models B5 and G5).

## References


Andersson, U. (2007). The contribution of working memory to children's mathematical word problem solving. *Applied Cognitive Psychology, 21*(9), 1201–1216. https://doi.org/10.1002/acp.1317.

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556–559. https://doi.org/10.1126/science.1736359.

Bae, Y. S., Chiang, H. M., & Hickson, L. (2015). Mathematical word problem solving ability of children with autism spectrum disorder and their typically developing peers. *Journal of Autism and Developmental Disorders, 45*, 2200–2208. https://doi.org/10.1007/s10803-015-2387-8.

Bentler, P. M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. *Psychological Bulletin, 106*(2), 315–317. https://doi.org/10.1037/0033-2909.106.2.315.

Berberoglu, G. (1995). Differential item functioning (DIF) analysis of computation, word problem and geometry questions across gender and SES groups. *Studies in Educational Evaluation, 21*(4), 439–456. https://doi.org/10.1016/0191-491X(95)00025-P.

Björn, P. M., Aunola, K., & Nurmi, J. E. (2016). Primary school text comprehension predicts mathematical word problem-solving skills in secondary school. *Educational Psychology, 36*(2), 362–377. https://doi.org/10.1080/01443410.2014.992392.

Carr, M., & Jessup, D. L. (1997). Gender differences in first-grade mathematics strategy use: Social and metacognitive influences. *Journal of Educational Psychology, 89*(2), 318328. https://doi.org/10.1037/0022-0663.89.2.318.

Casey, B. M., Lombardi, C. M., Pollock, A., Fineman, B., & Pezaris, E. (2017). Girls' spatial skills and arithmetic strategies in first grade as predictors of fifth-grade analytical math reasoning. *Journal of Cognition and Development,, 18*(5), 530–555. https://doi.org/10.1080/15248372.2017.1363044.

Chipman, S. F., Marshall, S. P., & Scott, P. A. (1991). Content effects on word problem performance: A possible source of test bias? *American Educational Research Journal, 28*(4), 897–915. https://doi.org/10.3102/00028312028004897.

Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review, 28*(4), 415–427. https://doi.org/10.1016/j.econedurev.2008.09.003.

Cohen, A. S., Gregg, N., & Deng, M. (2005). The role of extended time and item content on a high-stakes mathematics test. *Learning Disabilities Research & Practice, 20*(4), 225–233. https://doi.org/10.1111/j.1540-5826.2005.00138.x.

Cole, N. S. (1997). *The ETS gender study: How females and males perform in educational settings.* Princeton, NJ: Educational Testing Service.

Common Core State Standards Initiative (2010). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects.* Washington, DC: CCSSO & National Governors Association.

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20*(4), 405–438. https://doi.org/10.1016/0010-0285(88)90011-4.

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology, 6*, 348. https://doi.org/10.3389/fpsyg.2015.00348.

Davis, H., & Carr, M. (2002). Gender differences in mathematics strategy use: The influence of temperament. *Learning and Individual Differences, 13*(1), 83–95. https://doi.org/10.1016/S1041-6080(02)00063-8.

De Corte, E., & Verschaffel, L. (1987). The effect of semantic structure on first graders' strategies for solving addition and subtraction word problems. *Journal for Research in Mathematics Education,* 363–381.

Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions, 8*(1), 33. https://doi.org/10.1186/1744-9081-8-33.

Donlan, C., Cowan, R., Newton, E. J., & Lloyd, D. (2007). The role of language in mathematical development: Evidence from children with specific language impairments. *Cognition, 103*(1), 23–33. https://doi.org/10.1016/j.cognition.2006.02.007.

Dunbar, S., Welch, C., Hoover, H. D., Forsyth, R. A., Frisbie, D. A., & Ansley, T. N. (2015). *Iowa assessments research and development guide.* Iowa City, IA: University of Iowa.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention.* New York, NY: Guilford Press.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*(3), 493–513. https://doi.org/10.1037/0022-0663.97.3.493.

Fuchs, L. S., & Fuchs, D. (2002). Mathematical problem-solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities, 35*(6), 564–574. https://doi.org/10.1177/00222194020350060701.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71*(1), 7–21. https://doi.org/10.1177/001440290407100101.

Fuchs, L. S., Fuchs, D., Compton, D. L., Hamlett, C. L., & Wang, A. Y. (2015). Is word-problem solving a form of text comprehension? *Scientific Studies of Reading, 19*(3), 204–223. https://doi.org/10.1080/10888438.2015.1005745.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., ... 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. https://doi.org/10.1037/0022-0663.98.1.29.

Fuchs, L. S., Fuchs, D., Stuebing, K., Fletcher, J. M., Hamlett, C. L., & Lambert, W. (2008a). Problem solving and computational skill: Are they shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology, 100*(1), 30–47. https://doi.org/10.1037/0022-0663.100.1.30.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., ... Schatschneider, C. (2010a). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental psychology, 46*(6), 1731–1746. https://doi.org/10.1037/a0020662.

Fuchs, L.S., Gilbert, J.K., Fuchs, D., Seethaler, P.M., & Martin, B. (2018). Text comprehension and oral language as predictors of word-problem solving: Insights into word-problem solving as a form of text comprehension. Scientific Studies of Reading, 22, 152-166. First published online 21 Nov 2017; PMID 29643723; PMCID: PMC5891146; https://doi.org/10.1080/10888438.2017.1398259.

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 prealgebraic knowledge. *Developmental Psychology, 52*(12), 2085. https://doi.org/10.1037/dev0000227.

Fuchs, L. S., Powell, S. R., Cirino, P. T., Schumacher, R. F., Marrin, S., Hamlett, C. L., ... Changas, P. C. (2014). Does calculation or word-problem instruction provide a stronger route to prealgebraic knowledge? *Journal of Educational Psychology, 106*(4), 990–1006. https://doi.org/10.1037/a0036793.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., ... Zumeta, R. O. (2009). Remediating number combination and word problem deficits among students with mathematics difficulties: A randomized control trial. *Journal of Educational Psychology, 101*(3), 561–576. https://doi.org/10.1037/a0014701.

Fuchs, L. S., Hamlett, C. L., & Powell, S. R. (2003). Math Fact Fluency and Double-Digit Addition and Subtraction Tests. Available from LS Fuchs, 228 Peabody, Vanderbilt University, Nashville, TN 37203.

Fuchs, L. S., & Seethaler, P. M. (2008). Find X, number sentences, and Vanderbilt story problems. Nashville, TN 37203: Available from L. S. Fuchs, 228 Peabody, Vanderbilt University.

Fuchs, L. S., Seethaler, P. M., Powell, S. R., Fuchs, D., Hamlett, C. L., & Fletcher, J. M. (2008b). Effects of preventative tutoring on the mathematical problem solving of third-grade students with math and reading difficulties. *Exceptional Children, 74*(2), 155–173. https://doi.org/10.1177/001440290807400202.

Fuchs, L. S., Zumeta, R. O., Schumacher, R. F., Powell, S. R., Seethaler, P. M., Hamlett, C. L., & Fuchs, D. (2010b). The effects of schema-broadening instruction on second graders' word-problem performance and their ability to represent word problems with algebraic equations: A randomized control study. *The Elementary School Journal, 110*(4), 440–463. https://doi.org/10.1086/651191.

Geary, D. C. (1994). *Children's mathematical development: Research and practical applications.* Washington, DC: American Psychological Association.

Geary, D. C., Saults, S. J., Liu, F., & Hoard, M. K. (2000). Sex differences in spatial

cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology, 77*(4), 337–353. https://doi.org/10.1006/jecp.2000.2594.
Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*, 41–45.
Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87*(1), 18–32. https://doi.org/10.1037/0022-0663.87.1.18.
Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1993). *Iowa tests of basic skills (Levels 9-14)*. Chicago: Riverside Publishing.
Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127–160.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. https://doi.org/10.1080/10705519909540118.
IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.
Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology, 96*(4), 284–309. https://doi.org/10.1016/j.jecp.2006.09.001.
Jõgi, A. L., & Kikas, E. (2016). Calculation and word problem-solving skills in primary grades–Impact of cognitive abilities and longitudinal interrelations with task-persistent behaviour. *British Journal of Educational Psychology, 86*(2), 165–181. https://doi.org/10.1111/bjep.12096.
Jordan, N. C., & Hanich, L. (2000). Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities, 33*, 567–578. https://doi.org/10.1177/002221940003300605.
Jordan, N. C., Levine, S. C., & Huttenlocher, J. (1995). Calculation abilities in young children with different patterns of cognitive functioning. *Journal of Learning Disabilities, 28*(1), 53–64. https://doi.org/10.1177/002221949502800109.
Kail, R., & Hall, L. K. (1999). Sources of developmental change in children's word-problem performance. *Journal of Educational Psychology, 91*(4), 660–668. https://doi.org/10.1037/0022-0663.91.4.660.
Kendeou, P., Van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*(4), 765–778. https://doi.org/10.1037/a0015956.
Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*(1), 109–129. https://doi.org/10.1037/0033-295X.92.1.109.
Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology, 100*(2), 310–321. https://doi.org/10.1037/0022-0663.100.2.310.
Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3th Ed.). Guilford: New York, NY.
Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th Ed.). Guilford: New York, NY.
Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences, 13*(2), 129–164. https://doi.org/10.1207/s15327809jls1302_1.
Koedinger, K. R., & Tabachneck, H. J. (1994). *Two strategies are better than one: Multiple strategy use in word problem solving.* New Orleans, LA: Annual meeting of the American Educational Research Association.
Kong, J. E., & Orosco, M. J. (2016). Word-problem-solving strategy for minority students at risk for math difficulties. *Learning Disability Quarterly, 39*(3), 171–181. https://doi.org/10.1177/0731948715607347.
Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cerebral Cortex, 12*(5), 477–485. https://doi.org/10.1093/cercor/12.5.477.
Kyttälä, M., & Björn, P. M. (2014). The role of literacy skills in adolescents' mathematics word problem performance: Controlling for visuo-spatial ability and mathematics anxiety. *Learning and Individual Differences, 29*, 59–66. https://doi.org/10.1016/j.lindif.2013.10.010.
Laski, E. V., Casey, B. M., Qingyi, Y., Dulaney, A., Heyman, M., & Dearing, E. (2013). Spatial skills as a predictor of first grade girls' use of higher-level arithmetic strategies. *Learning and Individual Differences, 23*, 123–130. https://doi.org/10.1016/j.lindif.2012.08.001.
Lewis, K. E., & Fisher, M. B. (2016). Taking stock of 40 years of research on mathematical learning disability: Methodological issues and future directions. *Journal for Research in Mathematics Education, 47*(4), 338–371. https://doi.org/10.5951/jresematheduc.47.4.0338.
Loveless, T. (2015). The gender gap in reading. Brookings. Retrieved from: https://www.brookings.edu/research/the-gender-gap-in-reading/.
Low, R., & Over, R. (1993). Gender differences in solution of algebraic word problems containing irrelevant information. *Journal of Educational Psychology, 85*(2), 331–339. https://doi.org/10.1037/0022-0663.85.2.331.
MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149. https://doi.org/10.1037/1082-989X.1.2.130.
Magnuson, K., Duncan, G. J., Lee, K. T., & Metzger, M. W. (2016). Early school adjustment and educational attainment. *American Educational Research Journal, 53*(4), 1198–1228. https://doi.org/10.3102/0002831216634658.
Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x.
Marshall, S. P., & Smith, J. D. (1987). Sex differences in learning mathematics: A longitudinal study with item and error analyses. *Journal of Educational Psychology, 79*(4), 372–383. https://doi.org/10.1037/0022-0663.79.4.372.
Mathematics Learning Study Committee, & National Research Council. (2001). Adding it up: Helping children learn mathematics. Washington, D.C.: National Academies Press.
Mayer, R. E. (2013). Problem solving. In D. Reisberg (Ed.). *The Oxford handbook of cognitive psychology* (pp. 769–778). New York, NY: Oxford University Press.
Mercer, C. (1997). *Students with learning disabilities.* Upper Saddle River, NJ.: Prentice Hall.
Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105*(3), 430–445. https://doi.org/10.1037/0033-2909.105.3.430.
Murnane, R. J., Willett, J. B., Braatz, M. J., & Duhaldeborde, Y. (2001). Do different dimensions of male high school students' skills predict labor market success a decade later? Evidence from the NLSY. *Economics of Education Review, 20*(4), 311–320. https://doi.org/10.1016/S0272-7757(00)00056-X.
Muth, K. D. (1984). Solving arithmetic word problems: Role of reading and computational skills. *Journal of Educational Psychology, 76*(2), 205–210. https://doi.org/10.1037/0022-0663.76.2.205.
Muthén, L. K., & Muthén, B. O. (1998-2013). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.
National Mathematics Advisory Panel. (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. US Department of Education.
National Center for Education Statistics (2017). *NAEP 2017 reading: A report card for the nation and the states.* Washington, D.C.: U.S. Department of Education, Institute of Education Sciences.
Passolunghi, M. C., & Mammarella, I. C. (2010). Spatial and visual working memory ability in children with difficulties in arithmetic word problem solving. *European Journal of Cognitive Psychology, 22*(6), 944–963. https://doi.org/10.1080/09541440903091127.
Passolunghi, M. C., & Siegel, L. S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *Journal of Experimental Child Psychology, 80*(1), 44–57. https://doi.org/10.1006/jecp.2000.2626.
Petscher, Y., & Kim, Y. S. (2011). The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal of School Psychology, 49*(1), 107–129. https://doi.org/10.1016/j.jsp.2010.09.004.
Pickering, S., & Gathercole, S. (2001). *Working memory test battery for children.* London: The Psychological Corporation.
Pina, V., Fuentes, L. J., Castillo, A., & Diamantopoulou, S. (2014). Disentangling the effects of working memory, language, parental education, and non-verbal intelligence on children's mathematical abilities. *Frontiers in Psychology, 5*, 415. https://doi.org/10.3389/fpsyg.2014.00415.
Powell, S. R., Fuchs, L. S., Fuchs, D., Cirino, P. T., & Fletcher, J. M. (2009). Do word-problem features differentially affect problem difficulty as a function of students' mathematics difficulty with and without reading difficulty? *Journal of Learning Disabilities, 42*(2), 99–110. https://doi.org/10.1177/0022219408326211.
Purpura, D. J., Hume, L. E., Sims, D. M., & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology, 110*(4), 647–658. https://doi.org/10.1016/j.jecp.2011.07.004.
Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*(1), 144–161. https://doi.org/10.1037/0022-0663.88.1.144.
Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57*(1), 55–71. https://doi.org/10.1111/0022-4537.00201.
Rivera-Batiz, F. L. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *Journal of Human Resources, 313–328*. https://doi.org/10.2307/145737.
Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *The Journal of Educational Research, 95*(5), 259–272. https://doi.org/10.1080/00220670209596600.
Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. J., & Marchant, H. (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology, 24*(3), 181–266. https://doi.org/10.1006/ceps.1999.1004.
Seethaler, P. M., Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). Predicting first graders' development of calculation versus word-problem performance: The role of dynamic assessment. *Journal of Educational Psychology, 104*(1), 224–234. https://doi.org/10.1037/a0024968.
Schrank, F. A., McGrew, K. S., & Woodcock, R. W. (2001). Technical Abstract (Woodcock-Johnson III Assessment Service Bulletin No. 2). Itasca, IL: Riverside Publishing.
Staub, F., & Reusser, K. (1995). The role of presentational structures in understanding and solving mathematical word problems. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.). *Discourse comprehension* (pp. 285–305). Hillsdale, NJ: Erlbaum.
Steele, M. M. (2002). Strategies for helping students who have learning disabilities in mathematics. *Mathematics Teaching in the Middle School, 8*(3), 140–143.
Swanson, H. L. (2004). Working memory and phonological processing as predictors of children's mathematical problem solving at different ages. *Memory & Cognition, 32*(4), 648–661. https://doi.org/10.3758/BF03195856.
Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 96*(3), 471–491. https://doi.org/10.1037/0022-0663.96.3.471.

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. https://doi.org/10.1037/0022-0663.100.2.343.

Tajika, H., Nakatsu, N., Nozaki, H., Neumann, E., & Maruno, S. (2007). Effects of self--explanation as a metacognitive strategy for solving mathematical word problems. Japanese Psychological Research, 49(3), 222-233. doi: http://doi.org/10.1111./j.1468-5884.2007.00349.x.

Thevenot, C., Devidal, M., Barrouillet, P., & Fayol, M. (2007). Why does placing the question before an arithmetic word problem improve performance? A situation model account. *The Quarterly Journal of Experimental Psychology, 60*(1), 43–56. https://doi.org/10.1080/17470210600587927.

Tolar, T. D., Fuchs, L., Cirino, P. T., Fuchs, D., Hamlett, C. L., & Fletcher, J. M. (2012). Predicting development of mathematical word problem solving across the intermediate grades. *Journal of Educational Psychology, 104*(4), 1083–1093. https://doi.org/10.1037/a0029020.

Träf, U., & Samuelsson, J. (2013). An analysis of errors in multi-digit arithmetic and arithmetic word problem solving in children with mathematical difficulties. *Special Education, 1*, 121–132.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems.* Lisse, The Netherlands: Swets & Zeitlinger.

Vilenius-Tuohimaa, P. M., Aunola, K., & Nurmi, J. E. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology, 28*(4), 409–426. https://doi.org/10.1080/01443410701708228.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta analysis. *Psychological Bulletin, 140*(4), 1174. https://doi.org/10.1037/a0036620.

Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles, 41*(3), 219–240.

Wang, A. Y., Fuchs, L. S., & Fuchs, D. (2016). Cognitive and linguistic predictors of mathematical word problems with and without irrelevant information. *Learning and Individual Differences, 52*, 79–87. https://doi.org/10.1016/j.lindif.2016.10.015.

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence.* San Antonio, TX: The Psychological Corporation.

Wilson, W. S. (2009). Elementary school mathematics priorities. *AASA Journal of Scholarship & Practice, 6*(1), 40–49.

Woodcock, R. W. (1997). Woodcock diagnostic reading battery. Itasca, IL: Riverside. Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III. Itasca, IL: Riverside.

Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock–Johnson III Achievement Battery*. Itasca, IL: Riverside Publishing.

Xin, Y. P., Jitendra, A. K., & Deatline-Buchman, A. (2005). Effects of mathematical word problem-solving instruction on middle school students with learning problems. *The Journal of Special Education, 39*(3), 181–192. https://doi.org/10.1177/00224669050390030501.

Zheng, X., Swanson, H. L., & Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of Experimental Child Psychology, 110*(4), 481–498. https://doi.org/10.1016/j.jecp.2011.06.001, 749-750.

Zhu, J. (1999). *WASI manual.* San Antonio: The Psychological Corporation.