

Growth in Working Memory and Mathematical Problem Solving in Children at Risk and Not at Risk for Serious Math Difficulties

H. Lee Swanson
University of California, Riverside

Olga Jerman
Frostig School

Xinhua Zheng
University of California, Riverside

The influence of cognitive growth in working memory (WM) on mathematical problem solution accuracy was examined in elementary school children ($N = 353$) at risk and not at risk for serious math problem solving difficulties. A battery of tests was administered that assessed problem solving, achievement, and cognitive processing (WM, inhibition, naming speed, phonological coding) in children in 1st, 2nd, and 3rd grade across 3 testing waves. The results were that (a) children identified as at risk for serious math problem solving difficulties in Wave 1 showed less growth rate and lower levels of performance on cognitive measures than did children not at risk; (b) fluid intelligence and 2 components of WM (central executive, visual-spatial sketchpad) in Wave 1 (Year 1) predicted Wave 3 word problem solving solution accuracy; and (c) growth in the central executive and phonological storage component of WM was related to growth in solution accuracy. The results support the notion that growth in WM is an important predictor of children's problem solving beyond the contribution of reading, calculation skills, and individual differences in phonological processing, inhibition, and processing speed.

Keywords: working memory, math disabilities, problem solving, phonological processing, executive processing

Word problems are an important part of mathematics programs in elementary schools. This is because word problems help students apply formal mathematical knowledge and skills to real world situations. Much of the evidence indicates that word problem performance improves as children gain greater ability in (a) understanding underlying arithmetic operations (e.g., Rasmussen & Bisanz, 2005), (b) distinguishing between types of word problems on a basis of mathematical operations (e.g., Fayol, Abdi, & Gombert, 1987; Rittle-Johnson, Siegler, & Alibali, 2001), and (c) an effective use of strategies (e.g., Geary, Hoard, Byrd-Craven, & Desoto, 2004; Siegler, 1988). Developmental changes in skills in mathematics do not provide a complete account, however, of

age-related changes in word problem solving. There is evidence that suggests the need for more general cognitive processes, that is, processes nonspecific to mathematics. For example, solving a word problem, such as "15 dolls are for sale. 7 dolls have hats. The dolls are large. How many dolls do not have hats?" involves the development of a variety of mental activities (Barrouillet & L epine, 2005). Children must access prestored information (e.g., 15 dolls), access the appropriate algorithm (15 minus 7), and apply problem solving process to control its execution (e.g., ignore the irrelevant information). Given the multistep nature of word problems, it seems plausible that working memory (WM) plays a major role in solution accuracy. WM is defined as a processing resource of limited capacity, involved in the preservation of information while simultaneously processing the same or other information (Baddeley & Logie, 1999; Engle, Tuholski, Laughlin, & Conway 1999; Miyake, 2001).

Although there are several models of WM, Baddeley's multi-component model has often been used to explore the role of WM on mathematical problem solving (e.g., Swanson & Beebe-Frankenberger, 2004). Baddeley and Logie (1999) described WM as a limited capacity central executive system that interacts with a set of two passive store systems used for temporary storage of different classes of information: the speech-based phonological loop and the visual-spatial sketchpad. The phonological loop is responsible for the temporary storage of verbal information; items are held within a phonological store of limited duration, and the items are maintained within the store via the process of articulation (inner vocalization). The visual-spatial sketchpad is responsible for the storage of visual-spatial information over brief periods and

H. Lee Swanson, Graduate School of Education, University of California, Riverside; Olga Jerman, Frostig School, Pasadena, California; Xinhua Zheng, Department of Educational Psychology, University of California, Riverside.

This article is based on a 3-year longitudinal study funded by the U.S. Department of Education, Cognition and Student Learning (USDE R305H020055), Institute of Education Sciences awarded to H. Lee Swanson. We thank Georgia Doukas, Diana Dowds, Rebecca Gregg, Krista Healy, Crystal Howard, James Lyons, Kelly Rosston, and Leilani S aez for data collection and/or task development; the Colton School District, Tri City Christian Schools, and the Frostig School; and Margaret Beebe-Frankenberger and Bev Hedin who directed and managed data collection and school schedules. This article does not necessarily reflect the views of the U.S. Department of Education or the school districts.

Correspondence concerning this article should be addressed to H. Lee Swanson, Graduate School of Education, University of California, Riverside, CA 92521. E-mail: lee.swanson@ucr.edu

plays a key role in the generation and manipulation of mental images. Both storage systems are in direct contact with the central executive system. The central executive system is considered to be primarily responsible for coordinating activity within the cognitive system but also devotes some of its resources to increase the amount of information that can be held in the two subsystems (Baddeley & Logie, 1999). A recent formulation of the model (Baddeley, 2000) also includes a temporary multimodal storage component called the *episodic buffer*.

Although WM is a fundamental component in many current theories of children's problem solving (in addition to learning, comprehension, and reasoning), no longitudinal studies (to our knowledge) have explicitly isolated those components of WM most directly related to growth in problem solving, especially in children at risk for word problem solving difficulties. Because WM is made up of three components (central executive, phonological loop, and visual-spatial sketchpad), the question arises as to whether WM as a whole operates on problem solving or whether a certain component is more important. Thus, although there is a strong interrelationship among these components, we sought in the present study to determine whether growth in certain components of WM is related to growth in word problem solving. In this study, we attempted to break out the contributions of WM by focusing on short-term memory (STM) measures, verbal WM measures, and visual-spatial WM measures. In elaborating on the distinction between STM and WM, Cowan (1995) emphasized the role of attentional processes. WM is depicted as a subset of items of information stored in STM that are in turn submitted to limited attentional control processing (see also Engle et al., 1999). This assumes that when the contents of STM are separated from WM what is left of WM is controlled attention or processing related to the central executive system (also referred to in this article as *executive processing* or *the central executive component of WM*). Consequently, to understand the impact of WM to problem solving in terms of executive processing the influence of STM must be partialled out.

Models of WM and Word Problem Solving

How might growth in WM mediate age-related and individual differences in word problem solving? We consider three models as an explanation of the role of WM in individual and age-related problem solving performance in children: one focuses on the child's knowledge base for arithmetical calculations and components of word problems; another focuses on the storage components of WM, primarily the phonological loop; and the third focuses on the central executive system. The models are not necessarily exclusive of one another (each process can contribute important variance to problem solving to some degree) but suggest that some processes are more important than others.

Knowledge Base

The first model considers whether age-related differences in the child's knowledge base play a major role in mediating the influence of WM on problem solving. Several capacity models suggest that WM represents an activated portion of declarative long-term memory (LTM) (J. R. Anderson, Reder, & Lebiere, 1996; Cantor & Engle, 1993). That is, WM capacity influences the amount of

resources available to activate knowledge (see Conway & Engle, 1994, for a review of this model). Baddeley and Logie (1999) stated that a major role of WM "is retrieval of stored long-term knowledge relevant to the tasks at hand, the manipulation and recombination of material allowing the interpretation of novel stimuli, and the discovery of novel information or the solution to problems" (p. 31). They further stated that "any increase in total storage capacity beyond that of a given slave system is achieved by accessing either long-term memory (LTM) or other subsystems" (p. 37). Thus, the influence of WM performance on problem solving is related to one's ability to accurately access information (e.g., appropriate algorithm) from LTM to solve the problem. More specifically, a word problem introduces information into WM. The contents of WM are then compared with possible action sequences (e.g., associative links) in LTM (Ericsson & Kintsch, 1995). When a match is found (recognized), the contents of WM are updated and used to generate a solution. This assumption is consistent with current models of problem solving that are based on "recognize-act" models of a cognitive processor (J. R. Anderson et al., 1996; Ericsson & Kintsch, 1995).

In the present study, we assessed whether the retrievability of contents in LTM such as math skills, as well as domain-specific knowledge of the propositions found in word problems, mediates WM and problem solving. These LTM propositions are related to accessing numerical, relational, question, and extraneous information, as well as accessing the appropriate operations and solution algorithms (Hegarty, Mayer, & Monk, 1995; Mayer & Hegarty, 1996; Swanson, Cooney, & Brock, 1993). In the present study, we assumed that the contribution of LTM could be tested by partialing its influence from the correlations between WM and problem solving. Clearly, we only sampled some aspects of LTM, but we assumed that the influence of WM on problem solution accuracy should be eliminated when measures related to LTM (e.g., knowledge of the propositions found in word problems) are partialled from the analysis.

Phonological System

The second model assumes that the individual and age-related influence of WM on children's mathematic problem solving is primarily moderated by STM storage, the phonological system in particular. Because mathematical word problems are presented in a text format and the decoding and comprehension of text draws on the phonological system (see Baddeley, Gathercole, & Papagno, 1998, for a review), individual differences and age-related differences on problem solving tasks can be attributed to the phonological loop. Several studies assume that STM measures capture a subset of WM performance, the utilization and/or operation of the phonological loop (for a comprehensive review, see Dempster, 1985; Gathercole, 1998). This is because successful performance on STM measures draws on two major components of the phonological loop: a speech-based phonological input store and a rehearsal process (see Baddeley, 1986, for review). Research to date on STM indicates that young children rehearse less and perform more poorly on tasks requiring the short-term retention of order information when compared with older children (Ornstein, Naus, & Liberty, 1975), and children with math disabilities rehearse less when compared with children without math disabilities (e.g., see Geary, 2003, for a review). This suggests inefficient utilization of

the phonological rehearsal process (Henry & Millar, 1993). Because younger children and children with math disabilities have smaller digit spans than older children and children without math disabilities, it is possible they have basic inefficiencies in the storage of phonological input that influence higher level processing, such as comprehending and solving word problems. Developmental and individual differences in the phonological loop, therefore, might be expected to influence some aspects of problem solving, such as computing solutions to problems (Furst & Hitch, 2000). The phonological loop may be able to retain information of verbal form during ongoing calculations. Although not in the domain of math, per se, some studies suggest that simple short-term storage has a significant role in accounting for the relationship between WM and several cognitive abilities (Ackerman, Beier, & Boyle, 2005; Colom, Abad, Rebollo, & Shih, 2005; Colom, Flores-Mendoza, Quiroga, & Privado, 2005).

Thus, a simple version of this hypothesis states that individuals at risk for math disabilities and younger children are slower and/or less accurate at processing verbal information (numbers, letters) than average achieving children or older children, and this reduced processing on the participants' part underlies their poor WM and problem solving performance. These assumptions are consistent with a number of bottom-up models of higher order processing, such as comprehension, which view the primary task of executive processing as one of relaying the results of lower level linguistic analyses upward through the language system (Shankweiler & Crain, 1986). Phonologically analyzed information is transferred to WM storage, which in turn is then transferred (thus freeing storage for the next chunk of phonological information) upward through the processing system to promote online extraction of meaning. One of the possible reasons WM span increases as a function of age is because older children and children without math disabilities can name items more rapidly at recall than younger children or children with math disabilities. That is, increases in naming from early to the late childhood years are assumed to enhance the effectiveness of subvocal rehearsal processes and hence reduce the decay of memory items in the phonological store prior to output (Henry & Millar, 1993).

There are clear expectations in the aforementioned model: Individual and age-related changes in children's problem solving are related to the phonological loop. Mathematical proficiency follows automatically from improvements in phonological processing (i.e., because of improved storage and speed of processing information). Therefore, correlations between WM and problem solving should be significantly weakened if measures reflective of the phonological loop are partialled from the analysis. In other words, if age-related differences in problem solving performance and WM are moderated by the phonological system, then the relationship between problem solving and WM should be diminished when measures of the phonological system (e.g., phonological knowledge, speed, STM) are partialled from the analysis.

Central Executive System

The third model incorporates some of the assumptions of the first two models by viewing executive processing as: (a) providing resources to lower order (phonological system) skills and (b) accessing information from LTM. However, the model also views executive processes that are independent of those skills as playing

a major role in individual differences in mathematical problem solving. That is, although individual and age-related differences in problem solving accuracy are possibly related to the retrievability of contents in LTM (e.g., knowledge of specific mathematical relations, general problem solving strategies) and the phonological loop, other activities of the executive system may also underlie the influence of WM on solution accuracy (e.g., Swanson & Ashbaker, 2000; Swanson & Sachse-Lee, 2001). For example, several other cognitive activities (e.g., see Miyake, Friedman, Emerson, Witzki, & Howerter, 2000, for a review), such as controlling subsidiary memory systems, control of encoding and retrieval strategies, attention switching during manipulation of material held in the verbal and visual-spatial systems, and suppressing irrelevant information, have been assigned to the central executive system (Baddeley, 1996; Miyake et al., 2000; Oberauer, Suss, Wilhelm, & Wittman, 2003). Thus, the third model suggests that the central executive system contributes significant variance to individual and age-related differences in problem solving. Thus, in contrast to the aforementioned models that suggest increases in knowledge base and phonological processes play the dominant role in the mediating effects of WM on children's word problem solving, the present model assumes that executive processes also play a major role in mediating problem solving performance. There are clear expectations for this model. The influence of WM on measures of problem solving follows automatically with improvements in performance related to the executive component of WM.

Previous Individual and Cross-Sectional Studies

Some preliminary work has been done in addressing these three models. Studies by Swanson and colleagues (Swanson, 2004; Swanson & Beebe-Frankenberger, 2004; Swanson & Sachse-Lee, 2001) have shown that children with mathematical problem solving difficulties across different age groups perform poorly relative to controls on measures of WM. For example, Swanson and Sachse-Lee (2001) found with children ages 11 to 15 with and without math disabilities that phonological processing, verbal WM, and visual WM contributed unique variance to solution accuracy in word problems. This finding is in line with the current literature suggesting that the development of the phonological system plays an important part in accounting for individual differences in text processing. However, their results also showed that verbal WM processes played just as important a role as phonological processes (i.e., verbal WM reduced the contribution of the ability group contrast variable by approximately 63%) in accounting for the ability group differences in solution accuracy. Thus, there was weak support for the assumption that only the development of bottom-up processes (i.e., the phonological system) mediated deficits in WM processing and its influence on solution accuracy. In a follow-up analysis of Swanson and Beebe-Frankenberger's (2004) study, Swanson (2006) found that when residual variance related to STM was partialled out in the analysis, the component that best predicted problem solving was the central executive system of WM.

Unfortunately, none of the aforementioned studies have determined whether developmental increases in WM are related to developmental increases in problem solving. There have been some studies that have investigated cross-sectional differences in WM and word problem solving as a function of age (e.g., Swan-

son, 1999), however, no studies that we are aware of have investigated longitudinally whether changes in children's WM are related to changes in problem solving. Thus, despite studies showing that WM and problem solving improve as a function of age in cross-sectional studies (Swanson, 1999, 2003), the influence of growth in WM on problem solving in children has not been established. Identifying such a growth effect would be extremely valuable because it would assist in providing a parsimonious account of the ability group differences in WM found across a number of problem solving measures.

This study addressed the questions as to whether growth in problem solving is related to growth in WM and whether these changes vary as a function of children with and without serious mathematical problem solving disabilities in three elementary age groups and two ability groups (children at risk and not at risk for serious math difficulties). On the basis of the aforementioned models, we considered three possibilities: (a) Growth in the relationship between WM and problem solving is primarily mediated by task-specific knowledge and skills in mathematical calculation; (b) growth in WM and problem solving is primarily mediated by the phonological loop; or (c) growth in the executive component of WM, independent of the phonological system and resources activated in LTM, contributes unique variance to problem solving.

For this study, we assessed LTM using performance on measures of arithmetic calculation and recognition of problem solving components. We assessed the phonological system using performance on measures of naming speed, phonological awareness, and STM. We assessed executive processing using performance on WM tasks modeled after the format of Daneman and Carpenter's (1980) measure. The WM tasks were assumed to capture at least two factors of executive processing: susceptibility to interference and manipulation of capacity in the coordination of both processing and storage (e.g., Oberauer, 2002; Whitney, Arnett, Driver, & Budd, 2001). The approach used in this study (as well as in others, e.g., Engle, Cantor, & Carullo, 1992), to assess whether a particular system plays the major role in mediating differences in performance, was to remove statistically that system's influence from the analysis. In this study, the influence of the phonological system (e.g., naming speed, STM) or LTM was partialled via a hierarchical regression analysis between problem solving and WM. We reasoned that if WM and problem solving are primarily mediated by a phonological system and/or LTM, then the predictions of problem solving performance by performance on WM measures should be nonsignificant when measures related to the phonological system and LTM are entered (partialled) in the analysis. However, if growth in the central executive system (executive processing) mediates the relationship between WM and problem solving, then the correlations between these two variables will remain significant when measures of phonological processing and LTM are partialled from the analysis.

In summary, three questions directed our study.

1. Do children identified at risk or not at risk for serious math problem solving difficulties in Wave 1 vary on measures of growth in problem solving and WM across the three testing waves?

2. Is growth in WM related to growth in word problem solving?
3. Which components of WM are related to predictions of problem solving accuracy and are those predictions mediated by individual differences in knowledge base, phonological processing, and/or a central executive system?

Method

Participants

For the first wave of data collection, 353 children from Grades 1, 2, and 3 from a Southern California public school district and private school district participated in this study. Final selection of participants was determined by parent approval for participation and achievement scores. Of the 353 children selected, 167 were boys and 186 were girls. Gender representation was not significantly different among the three age groups, $\chi^2(2, N = 353) = 1.15, p > .05$. Ethnic representation of the sample was 163 Anglo, 147 Hispanic, 25 African American, 14 Asian, and 4 other (e.g., Native American and Vietnamese). The mean socioeconomic status of the sample was primarily middle class on the basis of parent education or occupation. However, the sample varied from low-middle class to upper middle class. Means and standard deviations for the variables used in this study for all testing waves are shown in Appendix A, B, and C.

The second wave of the testing 1 year later included 320 children, and the third wave of testing 2 years later included 302 children. The attrition of children who dropped out of the study was due to moving out of the school district. A comparison was made among achievement and cognitive scores, socioeconomic status, and parent income or occupation between children retained and those not retained in the study. In the Wave 1 sample, 134 children (68 boys, 66 girls) were classified at risk for serious math problem solving difficulties (SMD), and 219 children (99 boys and 120 girls) were not at risk. No significant differences emerged between the two ability groups in terms of ethnicity, $\chi^2(5, N = 353) = 8.67, p > .05$, or gender, $\chi^2(1, N = 353) = 0.06, p > .05$. With regard to attrition, however, the at-risk sample was reduced from 134 to 116 (64 boys and 52 girls) children, and the not-at-risk sample was reduced from 219 to 186 (94 boys and 92 girls) children in Wave 3. The comparison of the two groups at Wave 3 indicated that no significant differences emerged between the risk groups in terms of gender, $\chi^2(1, N = 302) = 0.61, p > .05$. Differences did emerge, however, between the risk groups in terms of ethnicity, $\chi^2(5, N = 302) = 12.03, p < .05$. Ethnic representation of the SMD ($n = 116$) sample in Wave 3 was 44 Anglo, 58 Hispanic, 7 African American, 4 Asian, and 3 other (e.g., Native American and Vietnamese). In contrast, ethnic representation for the children not at risk ($n = 186$) was 96 Anglo, 63 Hispanic, 14 African American, 12 Asian, and 1 other (Vietnamese). Thus, the at-risk group in Wave 3 had lower Anglo representation (only 38% of the sample) than the children not at risk (52% of the sample).

No bias was detected in child attrition except that proportionally more Hispanic children dropped out of the study. However, we did not find that gender interacted significantly with the performance on dependent measures (all $ps > .05$). Therefore, gender was not considered further in the analysis. The sample size and scores as a

function of age, ability group, and testing wave for each task are shown in Appendix A, B, and C.

Definition of Children at Risk for SMD

In this study, children at risk for SMD were defined as having normal intelligence on the basis of a measure of fluid intelligence (in this case, the Raven Colored Progressive Matrices test score ≥ 85 ; Raven, 1976) but with a mean performance below the 25th percentile (standard score of 90 or scaled score of 8) on norm-referenced measures related to (a) solving orally presented word problems and (b) digit naming fluency. The 25th percentile cutoff score on standardized achievement measures has been commonly used to identify children at risk (e.g., Fletcher et al., 1989) and, therefore, was used in this study. Classification of children at risk for SMD and not at risk for serious math difficulties was based on norm-referenced measures of computation on the Arithmetic subtest of the Wechsler Intelligence Scale for Children—Third Edition (WISC-III; Psychological Corporation, 1991) and Digit Naming Speed from the Comprehensive Test of Phonological Processing (CTOPP, Wagner, Torgesen, & Rashotte, 2000) described in the following section. Children who yielded an average scaled score less than or equal to 8 (average of both measures) were considered at risk for SMD. A scaled score of 8 was equivalent to a standard score of 90 or a percentile score of 25.

Our rationale for the above classification procedures was as follows. There are no general agreed upon criteria for defining children at risk for serious math difficulties in problem solving, especially in Grade 1 when the instruction is only beginning to address mathematical operations and learning to read. In addition, few studies have focused on math word problem solving difficulties in younger children, let alone identifying possible definitional parameters of risk. Because first graders were used in our sample in Wave 1, it was necessary in our classification of children at risk to control for the demands placed on reading and writing. In addition, because our focus was on examining problem solving skill and not arithmetic calculation, per se, for classification purposes it was necessary to rely on different measures than studies that define math disabilities by computation skill (i.e., mental computation of word problems versus paper and pencil computation of arithmetic problems). In contrast to the literature on math disabilities or reading disabilities, we assumed that children with SMD may not have skill difficulties related to arithmetic calculation or reading, per se, but may nevertheless have serious difficulties in coordinating arithmetic and language processes to solve a problem. For example, children we identified at risk had reading and math norm-referenced scores in the average range (see Appendix A, B, and C). Further, because we were interested in the reasoning processes related to problem solving, we selected a reliable measure that presented questions orally and placed problems in a verbal context (e.g., “If I have an candy bar and divide it in half, how many pieces do I have?”) rather than a written computational context (“ $1 + 1 = ?$ ”). Thus, we utilized the oral presentation of story problems as a criterion measure of SMD. Therefore, children who scored at or below a scaled score of 8 were selected as SMD. No doubt, this measure has been found to be a “complex mix of abilities, including quantitative reasoning . . . (a narrow ability subsumed by fluid reasoning), working memory, verbal comprehension and knowledge . . . the first choice for

interpretation of Arithmetic should likely be as a measure of fluid reasoning” (Keith, Fine, Taub, Reynolds, & Kranzler, 2006, p. 118). This interpretation was also supported in the WISC-III version (Keith & Wittam, 1997).

In selecting children with SMD at Wave 1, we also focused on a child’s verbal fluency with numbers during Wave 1 testing. We assumed that number naming speed may underlie children’s ability to automatically access arithmetic facts and procedures (Bull, Johnston, & Roy, 1999; Geary, Brown, & Samaranyake, 1991). We assumed that children who have quicker access to numbers (i.e., faster number fluency) would be less at risk for mental computation difficulties than children less fluent in number naming. This seemed reasonable to us based on Hitch and McAuley’s (1991) finding that children with math disabilities evidenced deficits in the speed of implicit counting. Further, speed of number naming has a parallel in the reading literature where both letter naming speed and phonological knowledge are assumed to underlie reading disabilities.

As will be shown in the Results section, the SMD classification of children in Wave 1 also differentiated children’s performance in Waves 2 and 3. However, it is important to note that a small number of children ($n = 15$; 13.16 % of the sample) did not stay at or below the 25th percentile (scale score of 8) across testing waves. Clearly, the problem solving and digit naming speed measure were not perfectly correlated and therefore the 25th percentile cutoff does not hold for both measures. For example, standard scores for the digit naming task were not stable over the three testing waves because children’s scale score moved from an 8 in Wave 1 to a 9 in Wave 3 (42% of the sample). We reran the analysis comparing at-risk and not-at-risk groups using only the WISC-III Math subtest as an indicator of ability (cutoff scores at or less than a scale score of 8). However, because the pattern in performance across measures was comparable, we used both measures to classify children at risk for SMD during Wave 1 testing.

Tasks and Materials

The battery of group and individually administered tasks is described below. Experimental tasks are described in more detail than published and standardized tasks. Tasks were divided into classification, criterion, and predictor variables. Cronbach’s alpha reliability coefficients for the sample were calculated for all scores across all testing waves.

Classification Measures

Fluid intelligence. Nonverbal intelligence was assessed by the Raven Colored Progressive Matrices test (Raven, 1976). We assumed that this measure tapped components of fluid intelligence in children (see Klauer, Willmes, & Phye, 2002; Stoner, 1982). Children were given a booklet with patterns displayed on each page, each pattern revealing a missing piece. For each pattern, six possible replacement pattern pieces were displayed. Children were required to circle the replacement piece that best completed the patterns. After the introduction of the first matrix, children completed their booklets at their own pace. Patterns progressively increased in difficulty. The dependent measure (range = 0–36) was the number of problems solved correctly, which yielded a standardized score ($M = 100$, $SD = 15$). The sample Cronbach’s

alpha coefficients on scores for Waves 1, 2, and 3 on this task were .87, .89, and .90, respectively.

Mental computation of word problems. This task was taken from the Arithmetic subtest of the WISC-III (Psychological Corporation, 1991). Each word problem was orally presented and was solved without paper or pencil. The dependent measure was the number of problems correct, which yielded a scaled score ($M = 10$, $SD = 2$). The sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 on this task were .74, .65, and .69, respectively.

Digit naming speed. The administration procedures followed those specified in the manual of the CTOPP (Wagner et al., 2000). For this task, the examiner presented participants with an array of 36 digits. Participants were required to name the digits as quickly as possible for each of two stimulus arrays containing 36 items, for a total of 72 items. The task administrator used a stopwatch to time participants on speed of naming. The sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 on this task were .92, .94, and .95, respectively.

Criterion Variables

Word problems—semantic structure varied. The purpose of this experimental measure was to assess mental problem solving as a function of variations in the semantic structure of a word problem (see Swanson & Beebe-Frankenberger, 2004, for details on this measure). Children were orally presented a problem and were asked to calculate the answer in their head. The word problems were derived from the work of Riley, Greeno, and Heller (1983), Kintsch and Greeno (1985), and Fayol et al. (1987). There are four sets of questions. Eight questions within each set were ordered by the difficulty of responses. The dependent measure was the number of problems solved correctly. The total possible number of questions correct was 32. The sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 for total scores on this task were .69, .54, and .65, respectively.

Word Problem Solving Components

Mathematical word problem solving processes. This experimental test assessed a child's ability to retrieve the propositions or components related to word problems (see Swanson & Beebe-Frankenberger, 2004, for details on this measure). The components assessed were derived from the earlier work of Mayer (see Mayer & Hegarty, 1996, for a review). Four booklets were administered that each contained three word problems and a series of multiple-choice questions. Problems were four sentences in length and contained two-assignment propositions (one relation, one question) and an extraneous proposition related to the solution. To control for reading problems, an examiner orally read each problem and all multiple-choice response options as the students followed along. Questions assessed the students' ability to correctly (a) identify the question proposition of each story problem, (b) identify the numbers in the propositions of each story problem, (c) identify the goals in the assignment propositions of each story problem, (d) correctly identify the operation, and (e) correctly identify the algorithm for each story problem. At the end of each booklet students were read a series of true–false questions. All statements were related to the extraneous propositions for each

story problem within the booklet. The total score possible for propositions related to question, number, goal, operations, algorithms, and true–false questions was 12. The sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 for total scores on this task were .97, .88, and .91, respectively.

Arithmetic Calculation

Arithmetic computation. The Arithmetic subtest from the Wide Range Achievement Test—Third Edition (WRAT-III; Wilkinson, 1993) and the Wechsler Individual Achievement Test (WIAT; Psychological Corporation, 1992) were administered. The sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 for the WRAT-III were .78, .84, and .71, respectively. For the WIAT, the sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 were .89, .86, and .76, respectively.

Computation fluency. This test was adapted from the Test of Computational Fluency (Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000). The adaptations required students to write answers within 2 min to 50 problems (25 on each page) of basic facts and algorithms. The basic facts and algorithms were problems matched to grade level (for Grades 1–5). The sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 for this task were .91, .97, and .98, respectively.

Predictor Variables

Phonological Knowledge Measures

Pseudoword reading tasks. The Pseudoword subtest was administered from the Test of Word Reading Efficiency (TOWRE; Wagner & Torgesen, 1999). The subtest required oral reading of a list of 120 pseudowords of increasing difficulty. The dependent measure was the number of words read correctly in 45 s. The sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 for this task were .94, .93, and .93, respectively.

Phonological deletion. The Elision subtest from the CTOPP (Wagner et al., 2000) was administered. The Elision subtest measures the ability to parse and synthesize phonemes. The child was asked to say a word and then say it again with a deleted part (e.g., "Say *popcorn*. Now say *popcorn* without saying *corn*"). The dependent measure was the number of items said correctly. The sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 for this task were .91, .88, and .89, respectively.

Reading

Word recognition. Word recognition was assessed by the Reading subtest of the WRAT-III (Wilkinson, 1993). The task provided a list of words of increasing difficulty. The dependent measure was the number of words read correctly. The sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 for this task were .91, .88, and .89, respectively.

Real word reading efficiency tasks. The Real Word Reading subtest was administered from the TOWRE (Wagner & Torgesen, 1999). The subtest required oral reading of a list of 120 real words of increasing difficulty. The dependent measure was the number of words read correctly in 45 s. The sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 for this task were .88, .96, and .93, respectively.

Reading comprehension. Reading comprehension was assessed by the Passage Comprehension subtest from the Woodcock Reading Mastery Test—Revised (WRMT-R; Woodcock, 1998). The dependent measure was the number of questions answered correctly. The sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 for this task were .93, .87, .84, respectively.

Naming Speed

Letter naming speed. The administration procedures followed those specified in CTOPP. Participants were required to name the letters as quickly as possible for each of two stimulus arrays containing 36 letters, for a total of 72 letters. The dependent measure was the total time to name both arrays of letters. The correlation between arrays in Forms A and B was .90. The sample Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 for this task were .91, .87, and .81, respectively.

STM Measures

Four measures of STM were administered: Forward and Backward Digit Span, Word Span, and Pseudoword Span. The digit subtest from the WISC-III was administered. The Forward Digit Span task required participants to recall and repeat in order sets of digits that were spoken by the examiner and that increased in number. The Backward Digit Span task from the WISC-III required participants to recall sets of digits in reverse order and was administered in the same manner as the Forward Digit Span task. The dependent measure was the highest set of items recalled in order (range = 0 to 6 for Backward Digit Span).¹ The Word Span and Pseudoword Span tasks were presented in the same manner as the Forward Digit Span measure. The Word Span task was previously used by Swanson, Ashbaker, and Lee (1996). The word stimuli are one- or two-syllable high-frequency words. Students are read lists of common but unrelated nouns and then are asked to recall the words. Word lists gradually increased in set size, from a minimum of two words to a maximum of eight. The Phonetic Memory task (Pseudoword Span task; Swanson & Berninger, 1995) uses strings of nonsense words (one syllable long), which are presented one at a time in sets of 2–6 nonwords. The dependent measure for all STM measures was the largest set of items retrieved in the correct serial order (range = 0–7). The Cronbach's alpha coefficients on scores for Waves 1, 2, and 3, respectively, scores for the Forward Digit Span were .78, .77, and .84; for the Backward Digit Span were .46, .62, and .48; for the Word Span were .70, .77, and .75; and for the Pseudoword Span were .58, .73, and .69.

WM Measures

The WM tasks in this study required children to hold increasingly complex information in memory while responding to a question about the task. The questions served as distracters to item recall because they reflected the recognition of targeted and closely related nontargeted items. A question was asked for each set of items and the tasks were discontinued if the question was answered incorrectly or if all items within a set could not be remembered. For this study, WM tasks were divided into those identified in the literature as tapping executive processing (Listen-

ing Span, Updating, Sentence–Digit task, Semantic Association task) and those assumed to tap the visual–spatial system (Visual–Matrix task, Mapping–Direction task). The complete description of the administration and scoring of the tasks is reported in Swanson (1995). Task descriptions follow.

Executive Processing

Listening sentence span. The children's adaptation (Swanson, 1992) of Daneman and Carpenter's (1980) Sentence Span Task was administered. This task required the presentation of groups of sentences, read aloud, for which children tried to simultaneously understand the sentence contents and to remember the last word of each sentence. The number of sentences in the group gradually increased from two to six. After each group was presented, the participant answered a question about a sentence and then was asked to recall the last word of each sentence. The dependent measure was the total number of correctly recalled word items up to the largest set of items (e.g., Set 1 contained two items, Set 2 contained three items, Set 3 contained four items, etc.) in which the process question was also answered correctly. Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 were .89, .89, and .83, respectively.

Semantic association task. The purpose of this task was to assess the participant's ability to organize sequences of words into abstract categories (Swanson, 1992, 1995). The participant was presented a set of words (one every 2 s per word), asked a discrimination question, and then asked to recall the words that "go together." For example, a child was first presented a list of the following words, *yellow, cake, red, cookies*, then was asked a discrimination question, "Did I say *red* or *orange*?" and then was asked to group the words by category (e.g., colors and food). In other words, the task required participants to transform information encoded serially into categories during the retrieval phase. The range of set difficulty was two categories with two words in each category in Set 1 and five categories of four words each in Set 7. Thus, the number of words in each set varied from 4 words in Set 1 to 20 words in Set 7. The dependent measure was the total number of correctly recalled words in which all items in the set were recalled and the process question for the set of words was answered correctly. Cronbach's alpha coefficients on scores for Waves 1, 2, and 3 were .84, .91, and .93, respectively.

Digit/sentence span. This task assesses the child's ability to remember numerical information embedded in a short sentence (Swanson, 1992, 1995). For example, Item 3 states, "Suppose

¹ There is some debate as to whether the backward Digit Span test better captures WM than STM (Gathercole, Pickering, Ambridge, & Wearing, 2004). As suggested by Colom, Abad, et al. (2005) and Engle et al. (1999), the numbers reversed task is assumed to be mainly a short-term processing capacity measure. Colom, Flores-Mendoza, et al. (2005) found forward and backward span measures to be similar measures of STM storage (see p. 1010 for discussion). Further, as stated by Engle et al. (1999) "Rosen and Engle (1997) showed that the backward and forward word task displayed similar effects of phonological similarity . . . suggesting that a simple transposition of order would be insufficient to move a task from the STM category to the WM category" (p. 314). In addition, Swanson, Mink, and Bocian (1999) found that with young children at risk for learning problems that both forward and backward digits loaded on the same factor as phonological processing (see Table 5).

somebody wanted to have you take them to the supermarket at 8-6-5-1 Elm Street.” The numbers are presented at 2-s intervals, followed by a process question (“What was the name of the street?”). The dependent measure was the total number of correctly recalled digits, in which all items in the set were recalled (set size ranged from 2 to 14 digits) and the process question for the set of digits was answered correctly. Cronbach’s alpha coefficients on scores for Waves 1, 2, and 3 were .88, .91, and .93, respectively.

Updating. This experimental updating task was adapted from Morris and Jones (1990). For this task, a series of one-digit numbers was presented that varied in set lengths of nine, seven, five, or three. No digit appeared twice in the same set. The examiner stated that the list may be either long or short and the participant should only remember the last three numbers in the same order as presented. Each digit was presented at approximately 1-s intervals. The four practice trials used list lengths of three, five, seven, and nine digits each, in random order. It was stressed that some of the lists of digits would be short, so they should not ignore any items. That is, to recall the last three digits in an unknown ($N = 3, 5, 7, 9$) series of digits, one must keep available the order of old information (previously presented digits), along with the order of newly presented digits. The dependent measure was the total number of lists correctly repeated (range = 0–16). Cronbach’s alpha coefficients on scores for Waves 1, 2, and 3 were .93, .89, and .89, respectively.

Visual–Spatial Sketchpad

Visual matrix task. The purpose of this task was to assess the ability of participants to remember visual sequences within a matrix (Swanson, 1992, 1995). Participants were presented a series of dots in a matrix and were allowed 5 s to study the matrix. The matrix was then removed and participants were asked, “Are there any dots in the first column?” After answering the discriminating question (by circling *y* for yes or *n* for no), students were asked to draw the dots they remembered seeing in the corresponding boxes of their blank matrix response booklets. The task difficulty ranged from a matrix of 4 squares and 2 dots (Set 1) to a matrix of 45 squares and 12 dots (Set 11). The dependent measure was the total number of items correct up to the highest set, in which the process question was answered correctly. Cronbach’s alpha coefficients on scores for Waves 1, 2, and 3 were .78, .79, and .84, respectively.

Mapping and directions. This task required a child to remember a sequence of directions on a map (Swanson, 1992, 1995). The experimenter presented a street map with dots (stop lights) connected by lines and arrows, which illustrated the direction a bicycle would go to follow a route through the city. After the map was removed, the child was asked a process question (“Were there any stop lights [dots] on the first street [column]?”). The child was then presented a blank matrix, on which to draw the street directions (lines and arrows) and stop lights (dots). The task difficulty on this subtest ranged from 2 dots and 3 lines (Set 1) to 20 dots and 23 lines (Set 9). The dependent measure was the total number of items correct up to the highest set, in which the process question was answered correctly. Cronbach’s alpha coefficients on scores for Waves 1, 2, and 3 were .78, .79, and .84, respectively.

Fluency and Inhibition Measures

Two tasks that we assume capture different aspects of controlled attention are fluency and random generation. Both tasks measure inhibition but emphasize different aspects. The fluency task requires individuals to spontaneously generate words in response to a category cue (e.g., generate animal names) or specific letter cue (generate words that begin with the letter *B*). These tasks have been associated with the executive processing that involves the controlled search-for words (e.g., see Rende, Ramsberger, & Miyake, 2002, for review). That is, participants are directed to activate needed information (animal names) while controlling the repetition of exemplars.

In the random generation procedure, on the other hand, participants are asked to keep track of the number of times that the items have been generated and to inhibit well-known sequences such as 1, 2, 3, 4 or a, b, c, d. This task differs somewhat from the fluency measure because participants must suppress rote or habitual responses (saying the letters of the alphabet in order) in order to quickly complete the task. Thus, during random generation the central executive system acts as a rate-limited filtering device that filters out habitual responses (see Towse, 1998, for a review of this measure).

Categorical fluency. This experimental measure was adapted from Harrison, Buxton, Husain, and Wise (2000). Children were given 60 s to generate as many names of animals as possible. Children were told, “I want to see how many animals you can name in a minute.” The dependent measure was the number of different words correctly stated within 60 s. Cronbach’s alpha coefficients on scores for Waves 1, 2, and 3 were .76, .76, and .75, respectively.

Letter fluency. This experimental measure was adapted from Harrison et al. (2000). Children were given 60 s to generate as many words as possible beginning with the letter *B*. The dependent measure was the number of words correctly stated in 60 s. Cronbach’s alpha coefficients on scores for Waves 1, 2, and 3 were .76, .76, and .75, respectively.

Random generation of letters and numbers. Each child was asked to write as quickly as possible numbers (or letters) first in sequential order to establish a baseline. Children were then asked to quickly write numbers (or letters) in a random nonsystematic order. Scoring included an index for randomness, information redundancy, and percentage of paired responses to assess the tendency of participants to suppress response repetitions. The measure of inhibition was calculated as the number of sequential letters or numbers divided by number of correctly unordered numbers or letters. It was our assumption that scores close to 1.0 would reflect a high and efficient ability to inhibit well-learned sequences. Cronbach’s alpha for the Letter Generation task for Waves 1, 2, and 3 scores were .52, .69, and .63, respectively, and for the Number Generation task were .75, .68, and .69, respectively.

Procedure

Five doctoral-level graduate students trained in test administration tested all participants in their schools in Waves 1, 2, and 3. One session of approximately 45–60 min was required for small group test administration and one session of 45–60 min for indi-

vidual administration for each wave. During the group testing session, data were obtained from the Raven Colored Progressive Matrices test, WIAT, WRAT-III, mathematical word problem solving process booklets, visual matrix test, and arithmetic calculation fluency. The remaining tasks were administered individually. We counterbalanced test administration to control for order effects. Task order was random across participants within each test administrator.

Statistical Method

Sequence of analysis. The results are organized into five parts. First, we determined whether the memory tasks fit a three-factor model (central executive, phonological loop, visual-spatial sketchpad). Thus, we computed a confirmatory factor model to determine the adequacy of fit to the data. In addition, we converted tasks that were conceptually related to single latent variables to reduce the sample-task ratio and to simplify the analyses.

Second, we compared performance of children at risk and not at risk for SMD across the testing waves. Because some initial differences emerged between the groups on measures of the Raven Colored Progressive Matrices test, scores from this test were used as a covariate in the multivariate analysis of covariance (MANCOVA). This analysis addressed the question as to whether children identified at risk or not at risk for SMD in Wave 1 vary on measures of growth in problem solving and WM across the three testing waves.

Third, we tested for convergence across the three age cohorts. Convergence consists of testing whether cohort groups tested at the same age overlap in performance and the parameters of growth are equal for each of the cohort groups (see E. R. Anderson, 1993, for a review). This test was necessary because we were using a cohort-sequential design. The design assumes that the parameters of growth (levels, slopes, and error terms) are invariant across the three age groups. Therefore, it was necessary to establish convergence across the three cohorts (children who started testing at Grade 1 vs. children who started testing at Grade 2 vs. children who started testing at Grade 3). To accomplish this, we conducted structural equation modeling using EQS 6.0 (Bentler, 2005) to determine whether the latent measures were invariant across the samples.

Fourth, after factor invariance was tested across the latent measures, a growth curve analysis was conducted using hierarchical linear modeling (HLM; Bryk & Raudenbush, 1992; Singer, 2002). We sought to isolate the components of WM that best related to word problem solving. We report our procedures for using HLM modeling in the next section. Overall, this analysis addressed the question as to whether growth in WM was related to growth in word problem solving. We also addressed the question as to whether ability groups varied in their rate of growth. The HLM method overcomes some of the limitation of our MANOVA because it does not assume that an equal number of repeated observations are taken for each individual or that all individuals were measured at the same time point. The model also allowed us to use random effects to model the continuous functions of age. Further, the HLM procedure does not require that missing data be ignored and provides a valid means to addressing standard errors. In contrast to traditional MANOVA repeated measures in which significance is tested against the residual error, the test of fixed

effects in mixed models is tested against the appropriate error terms as determined by the model specification.

The final section of the results focused on Wave 1 predictions of Wave 3 word problem solving. In this analysis, we were not interested in growth per se or the multilevel structure of the data but rather in how much of the variability in problem solving in Wave 3 was accounted for in Wave 1 data. Because the explained proportion of variance is somewhat problematic in HLM analysis (e.g., a large number of variance components, negative R^2 are possible, etc.; for discussion, see Snijders & Bosker, 1999, pp. 99–100), we used hierarchical regression models to simply isolate unique processes in Wave 1 that underlie word problem solving performance in Wave 3. This analysis allowed us to address the questions related to the three models (knowledge base, phonological system, central executive system) provided in the introduction. The key question addressed was whether performance related to three components of WM was related to predictions of problem solving accuracy and whether those predictions were mediated by individual differences in knowledge base, phonological processing, and/or executive processing.

A key assumption of this study is that partialing out the influence of STM from WM leaves residual variance related to controlled attention (Engle et al., 1999). Several studies have shown that WM and STM are distinct but highly related processes (e.g., see Heitz, Unsworth, & Engle, 2005, for a review). For example, Engle et al. (1999) investigated the relationship among measures of STM, WM, and fluid intelligence in adults. They found that STM and WM tasks loaded on two different factors. Although strong correlations emerged between the two factors ($r = .70$), they found that a two-factor model fit the data better than a one-factor model. Further, they found that by statistically controlling the variance between WM and STM factors, the residual variance related to the WM factor was significantly correlated with measures of intelligence. That is, they found a strong link between the latent measures of WM, but not STM, to fluid intelligence (e.g., also see Conway, Cowan, Bunting, Therriault, & Minkoff, 2002). They interpreted their findings as suggesting that the residual variance related to the WM factor corresponded to controlled attention of the central executive system. Thus, in our regression analysis, we entered measures of STM and WM simultaneously into the regression model to determine whether unique variance related to WM would predict problem solving accuracy.

HLM. As mentioned previously, we examined growth in problem solving using a multilevel framework referred to as HLM. We applied growth modeling to the study of intraindividual change in problem solving and WM over the 3-year period via the PROC MIXED program in SAS 9.1 (SAS Institute, 2003). The HLM procedure allowed us to determine both the average rate of change and individual variability in change over time. Age was the variable that represented the passage of time in our growth model. Because we had only three data points at best, our focus was on linear change and, therefore, a curvilinear relationship could not be reliably calculated. To interpret the results, we centered age at 9.7 (mean age at Wave 3), so that intercepts reflected the expected performance at that age. (It is important to note that slope remains the same whether the mean age is centered at Wave 1, 2, or 3. However, the correlation between the intercept and slope does vary as a function of different centering.) Thus, we centered our data on

the mean age at Wave 3 because we were interested in the final status of individual children after 3 years of math instruction.

Our growth model yielded parameter estimates that defined both the overall trajectory of the sample (fixed effects) and deviations in the overall trajectories (random effects). The model is expressed as:

$$y_{ij} = \beta_0 + \beta_1(\text{age}_{ij}) + U_{0j} + U_{1j}(\text{age}_{ij}) + R_{ij},$$

where y_{ij} is the dependent variable (problem solving) measured at time i in child j ; age_{ij} is the child j 's age at time i ; β_0 is the average intercept at 9.7; U_{0j} is the random intercept for child j ; U_{1j} is the random age slope for child j ; and R_{ij} is the residual for child j at time i . The between-children variance components, $\tau^2_0 = \text{Var}(U_{0j})$ and $\tau^2_1 = \text{Var}(U_{1j})$, reflect individual differences in level (intercept) and rate of change (slope), respectively. Thus, we estimated the association between the outcome (problem solving) and repeated measures of age across the 3-year time periods. We refer to this as our *unconditional growth model*.

Clearly, performance on word problems will be influenced by instructional procedures in the math classroom. At the start of the study, children were nested within different classrooms that reflect different instructional approaches of the math teacher (for the majority of participants, math instruction was in the homeroom class). Our analyses of random effects for intercept and slope included children nested within classroom (or math teacher). Thus, we modified the random effects (between-children effects) as $\tau^2_0 = \text{Var}(U_{0j}) \times (\text{Teacher or Math Class})$ and $\tau^2_1 = \text{Var}(U_{1j}) \times (\text{Teacher or Class})$ to reflect individual differences in level (intercept) and rate of change (slope), respectively. Classroom (math teacher) placement was not a continuous variable across the testing waves. For our analysis we selected Wave 1 math teachers for the nesting effect on the basis of our results from our administration of a National Science Foundation (2000) National Survey of Science and Mathematics Education. This survey showed the greatest amount of variance in instructional approaches among teachers occurred during Wave 1 when compared with other testing waves. Thus, our random effects for intercept and slope were nested within classroom at Year 1 (thus, we had randomly varying intercepts and growth rates within classroom). It is also important to note that we initially used a three-level model (Level 1 effects within person, Level 2 effects between persons nested within classroom, Level 3 effects between classrooms; see Singer, 2002, p. 167 for an example of this three-level growth model). However, because the random effects at Level 3 were not significant and less than 2% of the variance, these random effects were dropped from the analysis. Our approach was consistent with the model specification steps outlined in Snijders and Bosker's study (1999; see pp. 94–97), which called for excluding nonsignificant random effects.

In summary, our HLM (Level 1 within person, Level 2 between persons nested within classroom) included random effects from intercepts and slopes for children nested within the classroom of the math teacher. We used this unconditional growth model to examine problem solving performance as a function of the intercept and slope. For the fixed effect, the intercept provided information on the average level of the dependent variable at age 9.7 and the average rate of change across individuals. For the random effects, the intercept represented the variation (variance) around the intercept and the slope indicated whether there was variance related to change overtime. Significant random effects indicated that children differed in the level (intercepts) and/or rate of change (slopes).

After establishing our unconditional growth model, we tested whether entering the components of WM into the model explained any statistically significant associations obtained related to fixed effects and random effects. This is referred to as a *conditional model*. When one or more predictors are introduced into the conditional model, the reductions in the magnitude of the various components when compared with the unconditional model are analogous to effect sizes (Snijders & Bosker, 2003). This is similar to the use of R^2 in linear regression models. The primary distinction between a linear regression and HLM is that several R^2 values are relevant to HLM because there are several variance components.

Reliability estimates for the HLM model are based on the random effects. The random effects of the unconditional model represent the proportion of variance in that effect (i.e., parameter specific rather than error variance). If the random effect is significant and has satisfactory reliability, then it is appropriate to test whether additional variables can explain some of the variance in the unconditional model. Specifically, the conditional model predicted that the level and slope of problem solving performance would be associated with WM performance. To evaluate the compatibility of the data with our conditional model, we tested the significance of the model change. This was done by comparing the differences between the deviance values (i.e., the likelihood value for the correspondence between model and data) from the unconditional and conditional growth model. These are chi-square values, and the number of parameters added for the conditional model serves as degrees of freedom. In general, models with lower deviance fit better than models with higher deviance values. The deviance test can be used to perform a formal chi-square test in order to compare a model that adds certain effects versus a model that excludes them. For the present study, a significantly lower deviance score for the conditional model indicated that the conditional model showed a better fit to the data than the unconditional model.

We also compared children at risk for SMD and not at risk on parameter estimates of the intercept and slope. Because our classification of risk for SMD occurred at Wave 1, we centered age at Wave 3. Ability group differences have been established at Wave 1 (Swanson & Beebe-Frankenberger, 2004) and therefore it would be more informative about the stability of ability group differences by centering performance at Wave 3. Within this hierarchical model, we conducted several analyses comparing the risk and nonrisk groups identified in Wave 1 on intercept and growth scores across an array of measures.

Missingness. Missing data occur in longitudinal designs whenever an intended measurement is not obtained. We experienced some loss in our sample in Year 3 ($N = 353$ in Wave 1 and $N = 302$ in Wave 3). HLM allows for an incomplete data set, using data from all participants with at least two points. We used maximum likelihood procedures to determine parameter estimates because the maximum likelihood estimation procedure has several advantages over other missing data techniques (see Peugh & Enders, 2004, for discussion).

Results

The means and standard deviations for measures of intelligence, accuracy in recognizing problem solving components, problem solving solution accuracy, arithmetic calculation, phonological processing, reading, STM, WM, inhibition, and fluency for Waves 1, 2, and 3 are shown in Appendix A (starting with Wave 1 at

Grade 1), B (starting with Wave 1 at Grade 2), and C (starting with Wave 1 at Grade 3). Prior to testing our hypothesis, we considered normality of the data. Measures meet standard criteria for univariate normality with skewness for all measures less than 3 and kurtosis less than 4. Univariate outliers were defined as cases more than 3.5 standard deviations from the means. We examined multivariate outliers by calculating Mahalanobis' d^2 . None of the cases were deemed outliers.

Confirmatory Factor Analysis and Data Reduction

A critical assumption of this study was that three components were represented in our WM battery: executive processing, phonological loop, and visual-spatial sketchpad. In order to test this assumption, we ran a confirmatory factor analysis for Wave 1 data using the CALIS (Covariance Analysis and Linear Structural Equation) software program (SAS Institute, 2003) with the four WM tasks (Listening Span, Semantic Association, Digit-Sentence Span, Updating) loading on one factor, the four STM tasks (Forward Digit Span, Backward Digit Span, Pseudoword Span, Word Span) loading on a second factor, and the two visual WM tasks loading on a third factor (Visual Matrix, Mapping-Directions). The fit statistics were .91 for the comparative fit index (CFI; Bentler & Wu, 1995) and .05 for the root-mean-square residual error of approximation (Jöreskog & Sörbom, 1984), indicating a good fit to the data. All standardized parameters were significant at the .01 level. As we had no theoretical reasons for relying on a one-factor or two-factor model, the three-factor model was accepted as a good fit to the data for the total sample in Wave 1.

Because of the large number of variables, it was necessary to reduce two or more measures that were conceptually related to latent measures (factor scores). The weighting of these variables for Waves 1 and 2 are reported in Swanson's (2006) study. Task weightings for the latent measures used in regression analysis (to be discussed) are reported in Table 7. We used the SAS CALIS program to create factor scores for each set of measures with two or more variables. This procedure allowed us to calculate for each testing wave standardized beta weights. On the basis of the standardized loadings, we computed factor scores by multiplying the z score of the target variable by the factor loading weights for the total sample (see Nunnally & Bernstein, 1994, p. 518 for calculation procedures).² Factor scores were created for problem solving accuracy (WISC-III: Arithmetic; problem solving semantic structure varied), math calculation (WRAT: Math; WIAT: Math, Computation Fluency), word problem solving components (correctly identifying the word problem propositions-related questions, numerical information, goals, arithmetical operations, algorithm, and irrelevant information), reading (WRAT: Reading; WRMT: Reading Comprehension), phonological processing knowledge (TOWRE: Pseudowords, Elision), inhibition (random generation of letters and random generation of numbers), fluency (Phonological and Semantic Fluency), speed (rate naming speed of numbers and letters), phonological loop (Forward Digit, Backward Digit, Pseudoword Span, Real Word Span), visual-spatial sketchpad (Visual Matrix, Mapping/Directions), and executive processing (Listening Sentence Span, Digit/Sentence, Semantic Association, Updating).

Ability Group and Wave Comparisons

Figure 1 shows the latent measures as mean z scores for each measure used in the subsequent analysis as a function of the three

cohorts and two ability groups across the three testing waves. The figure shows the starting points by age for three cohorts at Wave 1 and follows their performance across the three testing waves. The age range was from 6 to 10 (Grades 1-5). Figure 1 also shows children identified at risk for SMD and those not at risk within each cohort. All figures were scaled to show a range of 2.5 standard deviations above and below a mean z score of 0 based on Wave 1 performance.

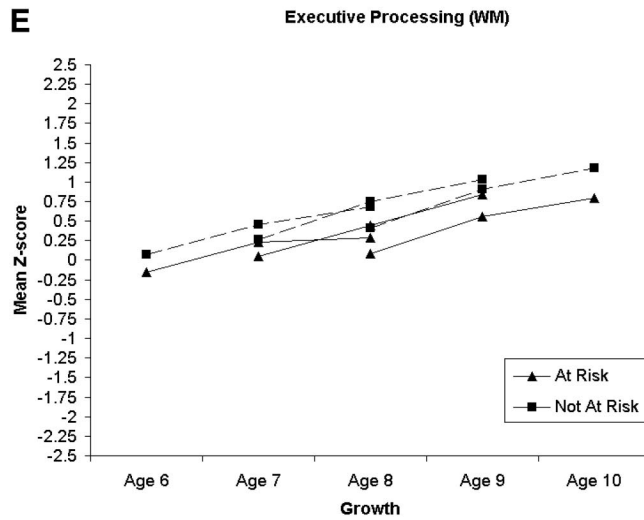
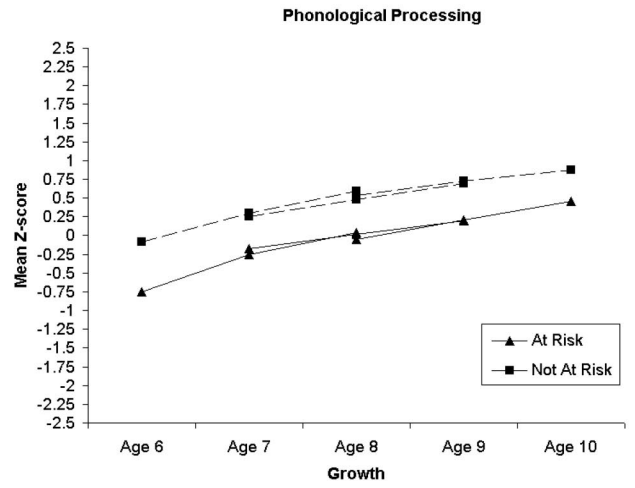
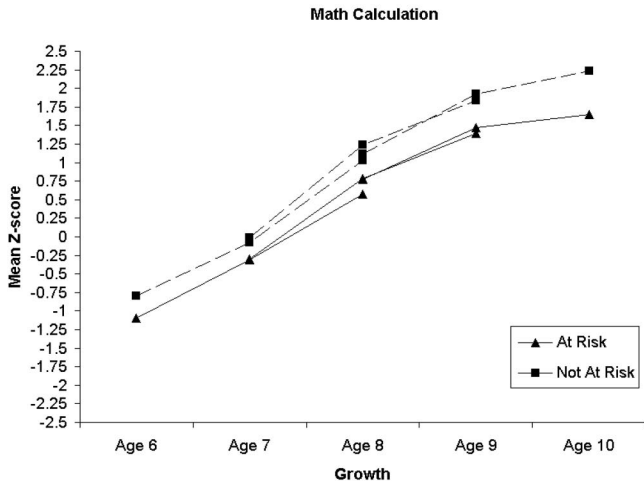
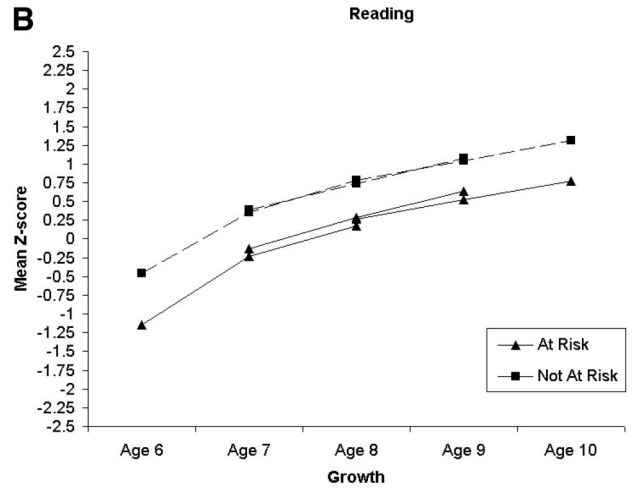
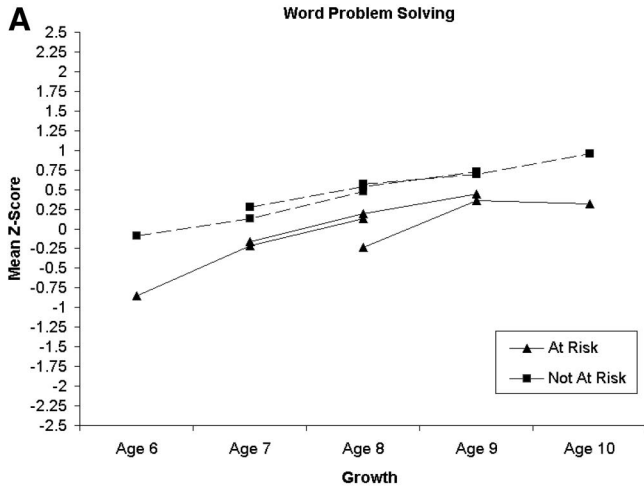
As shown in Figure 1, measures that showed the greatest increase from ages 6 to 10 (Grades 1-5) were math calculation, reading, and random generation. Measures showing the smallest rate of change ($< .25 SD$) were performance measures of fluency, phonological loop, and knowledge of word problem solving components. Across all measures, children identified at risk were lower in performance than those not at risk. It is important to note that problem solving from the WISC-III Arithmetic subtest and rapid naming of numbers from the CTOPP were used as classification variables and, therefore, it would be expected that differences emerged between these two groups in Wave 1 for problem solving and speed. However, we considered it an empirical question as to whether such differences at Wave 1 would be maintained in Wave 3.

The primary question addressed in this part of the analysis was whether performance of children at risk for SMD varied from

² Consistent with other investigations of growth (e.g., Wilson et al., 2002), we converted raw scores to z scores. All measures were scaled to have a mean of 0 and standard deviation of 1 at Wave 1. Wave 2 and 3 measures were z scored on the basis of the means and standard deviations of Wave 1. It was necessary to scale to z scores across the total sample so that all parameters were on the same metric, enabling meaningful comparisons for both age and time (Waves 1, 2, and 3) (see McGraw & Jöreskog, 1971, for a discussion). In addition, because of the number of variables, we created factor scores. These scores were calculated for each of the three testing waves. This was done for measurement purposes (different variables have different weightings on a construct) and also for practical reasons: Some constructs (e.g., WM, STM) included several tasks. One reviewer questioned whether our scaling of items would best be served using an item response theory model. The item response theory framework posits a log linear, rather than a linear, model to describe the relationship between observed item responses and the level of the underlying latent trait, θ . We agree with the reviewer's suggestion. However, as indicated by Nunnally and Bernstein (1994), for item response theory models "very large normative bases are required to implement all but the simplest and, therefore, sometimes unrealistic models using current estimation algorithms" (p. 396). In addition, as stated by Plewis (1996),

The feasibility of using vertical equating in the context of growth studies spanning a narrow range with different tests used at different ages has yet to be demonstrated. Certainly it is difficult to see how item response theory, which depends on fixing the mean and standard deviation of the ability distribution in order to estimate both item parameters . . . and an individual's ability, could be adapted to deal with growth without making arbitrary scaling assumptions. (p. 28)

Because we cannot address these issues with our current sample size (e.g., our sample is not large enough to establish anchor points), we attempted to place our values on the same measurement scale across the groups and investigate linear change. Converting data to z scores across the sample is appropriate when testing covariance structures (e.g., see Reise, Widaman, & Pugh, 1993, p. 557, for an example), as we have done in this study.



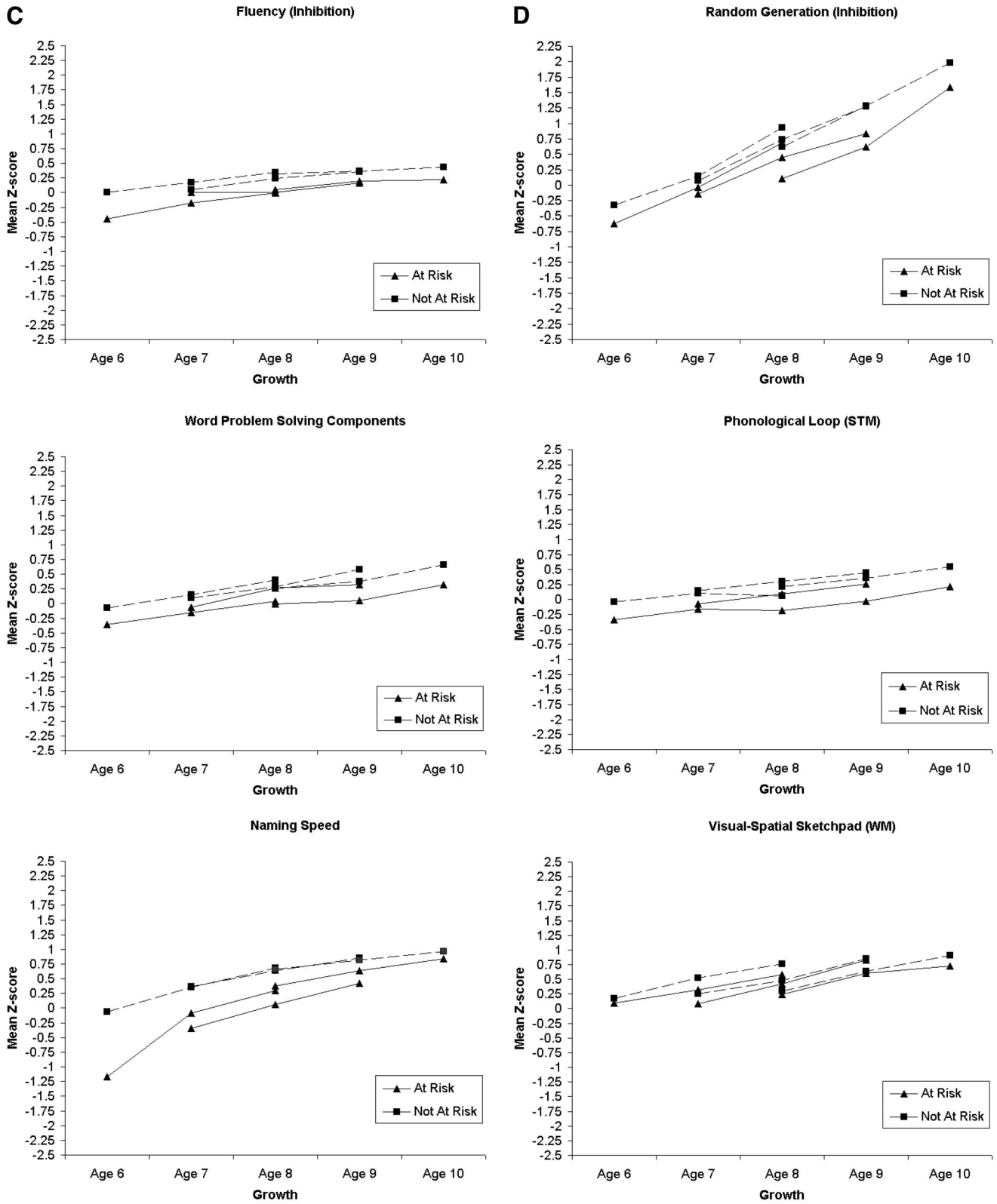


Figure 1 (opposite). Mean z scores for achievement and cognitive domains across testing waves as a function of age and ability group. STM = short-term memory; WM = working memory.

children not at risk on latent measures across testing waves. To address this question, we computed a 2 (risk group: children not at risk vs. SMD) × 11 (number of domains: problem solving accuracy, math calculation, word problem solving components, reading, phonological processing knowledge, inhibition, fluency, speed, phonological loop, visual–spatial sketchpad, and executive processing) × 3 (testing Waves 1, 2, and 3) repeated measures analysis of covariance (ANCOVA), with repeated measures on the last two factors. Although Raven scores were in the normal range for the at-risk group, an advantage was found for the not-at-risk group. Thus, Raven scores served as a covariate in the analysis. The covariate (Raven scores) was significant, $F(1, 266) = 10.01$, $MSE = 4.34$, $p < .001$, but the Ability Group × Covariate interaction was not significant, $F(1, 266) = 1.59$, $p = .20$, meeting the assumptions of a MANCOVA (homogeneity of the slopes). Thus, the covariate was judged to be reliable for the covariance analysis. Because there was a violation of sphericity (Mauchly’s criterion reported probability $< .0001$), we used the Greenhouse–Geisser and the Huynh–Feldt probability values. Because the samples of children not at risk and at risk for SMD

were matched on age, age was not considered as a covariate in this analysis. (The influence of age will be considered in the subsequent analyses.)

Omnibus F effects associated with the risk status, wave, and domain are reported below. The mean z scores of the latent measures are shown in Table 1 for Waves 1, 2, and 3. As shown in Table 1, the general pattern in the results was that children not at risk scored higher than those at risk for SMD, and scores were higher in testing Waves 2 and 3 than in Wave 1. Also provided in Table 1 is the difference in z scores between Wave 3 and Wave 1. As shown in Table 1, mean difference scores (Wave 3 minus Wave 1) greater than 1.0 standard deviation occurred for the measures of math calculation and inhibition. Although it is important to note that not all data can be used in this analysis (missing cells across waves are dropped), the results do show that differences between Wave 1 and Wave 3 were larger (z score differences $> .40$) for the children not at risk on measures of math calculation and knowledge of problem solving components. In contrast, difference scores were larger for the at-risk group on measures of problem solving accuracy and naming speed.

Table 1
Mean z Scores for Each Domain (Composite Score) as a Function of at Risk for Serious Math Difficulties (SMD; $n=104$) and Not at Risk for Serious Math Difficulties (NSMD; $n=166$)

Domain	Wave 1		Wave 2		Wave 3		Differences	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	LSM	<i>SD</i>
Problem solving								
SMD	−0.50	0.60	0.03	0.41	0.28	0.39	0.82	0.55
NSMD	0.32	0.37	0.50	0.40	0.72	0.48	0.40	0.45
Math								
SMD	−0.36	0.62	0.42	1.09	0.98	0.96	1.58	1.04
NSMD	0.21	0.70	1.30	1.41	1.78	0.92	2.12	1.06
Reading								
SMD	−0.57	0.84	0.07	0.70	0.43	0.68	1.04	0.50
NSMD	0.37	0.67	0.78	0.50	1.10	0.48	0.73	0.41
Phonological knowledge								
SMD	−0.45	0.62	−0.07	0.57	0.18	0.59	0.63	0.44
NSMD	0.30	0.63	0.54	0.54	0.74	0.50	0.44	0.45
Fluency								
SMD	−0.22	0.53	−0.02	0.44	0.09	0.38	0.32	0.44
NSMD	0.14	0.52	0.28	0.43	0.35	0.44	0.21	0.55
Problem solving components ^a								
SMD	−0.23	0.40	0.12	0.52	0.55	0.63	0.76	0.56
NSMD	0.13	0.34	0.59	0.53	1.17	0.61	1.05	0.52
Speed								
SMD	−0.67	1.17	0.08	0.64	0.44	0.53	1.12	0.84
NSMD	0.40	0.51	0.65	0.42	0.85	0.40	0.44	0.35
Inhibition								
SMD	−0.35	0.60	0.21	0.71	0.87	0.71	1.18	0.70
NSMD	0.26	0.73	0.83	0.85	1.54	0.99	1.34	0.72
Phonological loop								
SMD	−0.22	0.38	−0.17	0.40	−0.02	0.41	0.23	0.38
NSMD	0.12	0.39	0.15	0.44	0.38	0.47	0.25	0.44
Visual–spatial sketchpad								
SMD	−0.13	0.49	−0.08	0.60	0.24	0.72	0.39	0.78
NSMD	0.05	0.56	0.18	0.81	0.62	0.90	0.56	0.93
Executive working memory								
SMD	−0.19	0.39	0.06	0.55	0.37	0.61	0.59	0.57
NSMD	0.09	0.48	0.56	0.61	0.90	0.73	0.79	0.73

Note. LSM = least square means partialled for Raven scores.

^a Knowledge of problem solving components.

The repeated measures ANCOVA indicated that significant main effects emerged for ability group, $\lambda = .50$, $F(33, 235) = 6.92$, $p < .0001$, $\eta^2 = .50$; domain, $\lambda = .77$, $F(10, 258) = 7.62$, $p < .0001$, $\eta^2 = .23$; and testing wave, $\lambda = .83$, $F(2, 266) = 27.08$, $p < .0001$, $\eta^2 = .17$. Significant interactions emerged for the Group \times Domain interaction, $\lambda = .77$, $F(11, 259) = 7.61$, $p < .0001$, $\eta^2 = .23$; the Domain \times Wave interaction, $\lambda = .73$, $F(11, 248) = 4.41$, $p < .0001$, $\eta^2 = .27$; the Ability Group \times Wave interaction, $\lambda = .97$, $F(2, 266) = 3.24$, $p < .05$, $\eta^2 = .03$; and the Ability Group \times Domain \times Wave interaction, $\lambda = .66$, $F(11, 248) = 6.19$, $p < .001$, $\eta^2 = .34$.

A test of simple effects indicated that comparisons between at-risk and not-at-risk students were significant for each wave and domain. However, greater differences emerged between some domains than others. In general, a comparison of ability groups across domain (collapsed across time) indicated that the differences in favor of children not at risk, when compared with children at risk for SMD, were greater on all measures, except for the fluency and the storage components of WM (visual-spatial sketchpad and phonological loop). Further, when the difference scores between Wave 3 and Wave 1 served as a dependent measure (Wave 3 minus Wave 1), significant differences ($ps < .01$) in ANCOVAs were found between ability groups for all measures except for the measures of STM (phonological loop), fluency, visual spatial WM (visual-spatial sketchpad), and inhibition ($ps > .05$). As shown in Table 1, significant differences in scores emerged in favor of the nonrisk group on the remaining measures (all $ps < .01$).

In summary, the general trend was that children not at risk scored higher than those at risk, and higher difference scores emerged for academic domains (e.g., math calculation) when compared with cognitive domains (e.g., inhibition).

Convergence

A key assumption in growth modeling using a cohort sequent design is convergence. Convergence consists of testing whether cohort groups tested at the same age overlap in performance and

the parameters of growth (the levels, slopes, and error terms) are equal for each of the cohort groups (see E. R. Anderson, 1993, for a review). To address these assumptions, we used structural equation modeling (EQS 6.0; Bentler, 2005) to test whether the underlying linear growth processes (mean and variance of the slope and intercept factors) were invariant for the three cohort age groups (children in Wave 1 who started at Grade 1,2, or 3). In structural equation modeling, it is possible to test the invariance (equality) on all parameter estimates across multiple groups. We evaluated the invariance of various sets of parameters across the three age groups. The present analysis determined whether the intercepts and slopes across all individuals depicted a common line. A statistical test of convergence was done to determine whether the parameters (the intercepts, slopes, and error terms) were equal for each of the three cohort groups. The baseline model reflected configural invariance that tested whether similar patterns (not necessarily identical) emerged between the three groups across the three testing waves. Configural invariance requires only that the number of factors and factor loading patterns be the same across the groups (Byrne, 2006, see p. 233). A goodness-of-fit statistic related to this multigroup parameterization should be indicative of a well-fitting model. A well-fitting model has CFI at or above .90 (Bentler, 2005). As shown in Table 2, the CFIs ranged from .95 to 1.00 for the configural or baseline model, indicating that groups share a common structure and pattern across all measures. As shown in Table 2, three models tested the equality (invariance) between groups on various parameters. We tested for invariance across the three samples (a) when all parameters (intercepts, slopes, and errors) were constrained to be equal (Model 1), (b) when the intercept for each sample that included the same grade levels (e.g., Grade 3 in all three samples, Grade 2 in two samples) was constrained to be equal (Model 2, slope or errors were not constrained to be equal across groups), and (c) when the intercept and slope for the same grades in each sample were constrained to be equal (Model 3, errors are not constrained to be equal). Generally, it is argued that invariance holds if goodness-of-fit (CFI) to the model (e.g., Model 1, 2, or 3) is adequate and if there is no

Table 2
Invariance Tests Conducted Over Multiple Age Groups for a Cohort-Sequential Design

Domain	Configural model		Model 1: All constrained		Model 2: Intercepts constrained ^a		Model 3: Intercepts and slopes constrained	
	$\chi^2(3)$	CFI	$\chi^2(17)$	CFI	$\chi^2(9)$	CFI	$\chi^2(13)$	CFI
Word problems	10.49	.97	65.46	.91	19.92 ^a	.95	54.08	.94
Arithmetic calculations	87.59	.97	250.26	.71	103.98	.89	217.16	.77
Reading	51.54	.99	220.79	.90	68.37	.98	183.50	.93
Phonological processing	8.09	1.00	70.69	.97	16.65 ^a	.99	58.24	.98
Fluency	1.49	1.00	31.93	.92	12.03 ^a	.97	24.80	.98
WP components	45.58	.95	108.64	.90	53.82 ^a	.93	102.44	.88
Speed	25.46	.99	119.81	.94	38.43 ^a	.98	105.85	.95
Inhibition	5.96	1.00	28.96 ^a	1.00	6.28 ^a	1.00	21.79 ^a	1.00
Phonological loop	17.03	.99	38.24 ^a	.98	21.32 ^a	1.00	32.69 ^a	1.00
Sketchpad	12.33	.97	23.31 ^a	1.00	16.26 ^a	1.00	21.35 ^a	1.00
Executive working memory	3.92	1.00	34.07	.95	15.82 ^a	.94	31.76	.93

Note. CFI = comparative fit index; WP = word problem solving components accuracy score.

^a Converged.

statistically significant difference from that of the configural model (Byrne, 2006). Although several authors consider Model 1 unrealistic (see Byrne, 2006, pp. 244–247), we found no statistical difference between the configural model (baseline model) and Model 1 for performance related to inhibition, visual–spatial sketchpad, or the phonological loop. For example, a statistical test between the phonological loop and the configural model was not significant, $\Delta\chi^2(df = 14 [17 \text{ for Model 1}] - 3 [\text{for the configural model}]) = 21.21 (38.24-17.03), p > .05$. As shown in Table 2, Model 2 showed convergence (invariance) for all cognitive and word problem solving measures (all $\Delta\chi^2$ s $> .05$). This finding indicated that the intercept values for overlapping ages were statistically comparable. We were unable to find convergence, however, on measures of reading and math calculation suggesting that other variables (e.g., instruction) influenced cohort effects.

HLM

As noted above, cross-group equivalencies were excellent on all measures, except two (reading and math). Thus, some sources of invariance were unaccounted for across the cohort groups. Instead of uncovering each invariant parameter across the age groups, however, we decided to center our results on age, using HLM procedures. This is because the majority of findings related to intercepts and slopes are conditional on age (see Mehta & West, 2000), and, therefore, it was not reasonable to ignore age differences. As stated by Mehta and West (2000), conventional SEM models based on sample means and covariance structures use an “inappropriate within person scaling of the time variables yielding incorrect estimates of some of the random effect and increased complexity of interpretation of age effects” (p. 24). To correct this, for our next analysis (consistent with Mehta & West, 2000) we scaled age with respect to a common origin across all individuals to reflect deviations from a common age.

Unconditional Model

Table 3 shows our individual growth modeling to the study of intraindividual change in problem solving over the 3-year period. We estimated individual-level trajectories as well as overall sample-level trajectories. Table 3 shows an unconditional model, and Table 4 shows a conditional model. To interpret these tables, we first consider the unconditional model for the fixed effects in Table 3. The fixed effects model (which can be viewed as a one-way random effects analysis of variance model) provided the

estimates of the intercept (centered for the mean age at Wave 3) and growth. In this case, a mean z score of .58 was the predicted amount of solution accuracy at age 9.7 at Wave 3. The average unit of linear growth was .28. Hence, a 9.7-year-old ended Wave 3 with a z score of .58 and gained .28 units per testing session. Also shown are the t ratios (20.71 and 19.08) indicating that the parameter estimates were significantly greater than chance.

Also presented in Table 3 for the unconditional model are the model random effects for the total sample. The random effects include the intercept, slope, and within-child variance across testing waves. It is important to note that all students were exposed to variations in math instruction and, therefore, it was necessary to consider this variable in our calculation of random effects. We used as a representative variable the growth of students nested within classroom for Wave 1. Thus, our random effects (variance for intercept, slope) represented multiple observations of children (participants) nested within classrooms at Wave 1. As shown in Table 3, significant between-subjects variance emerged for the intercept (.18) and the slope (.03) estimates. Also included in the model was the within-person residual (.09). This estimate indicated that significant within-child variance emerged for problem solving accuracy across testing sessions.

Conditional Model-Group Effects

The conditional model in Table 4 shows a comparison of the two-risk groups identified at Wave 1 on problem solving accuracy. The difference between the two models is the addition of the classification (group) variable. For children at risk for SMD, the intercept (estimated z score for a child at 9.7 at the end of the study) for problem solving was .32, and linear growth rate was .41. These intercept and growth estimates were significantly better than chance. Likewise, for children not at risk, the intercept for word problem solving was .75 and the linear growth rate was .20. The intercept and linear growth estimates for the not-at-risk group were also significantly better than chance. In addition we determined whether the differences in estimated intercepts and growth rates varied significantly between the two groups. Ability group differences emerged in the estimated intercept values (.32 vs. .75), $F(1, 927) = 62.93, p < .0001$ (note df denominator reflects estimated missing values), and the slope values (.41 vs. .20), $F(1, 927) = 50.39, p < .0001$. Thus, although ability group differences would be expected at Wave 1 (because at-risk groups were separated by problem solving accuracy and rapid digit naming), the results show

Table 3
Growth on Word Problem Solving for the Unconditional Model

Effect	Parameter estimate	Variance estimate	SE	t	z
Fixed					
Intercept	.58		.02	20.71***	
Growth (linear)	.28		.01	19.08***	
Random (Participants \times Teachers)					
Intercept		.18	.02		8.23***
Growth (linear)		.03	.006		3.95***
Residual		.09	.007		12.23***

*** $p < .001$.

Table 4
Growth on Word Problem Solving for the Conditional Model

Effect	Group				F
	At risk for SMD		Not at risk for SMD		
	Parameter estimate	SE	Parameter estimate	SE	
Fixed					
Intercept	.32***	.04	.75***	.03	62.93***
Growth (linear)	.41***	.02	.20***	.02	50.39***
Random (Participants × Teachers)					
Subject	.14	.01	7.38***		
Growth	.02	.006	2.83**		
Residual	.08	.007	12.25***		

Note. SMD = serious math difficulties.
 ** $p < .01$. *** $p < .001$.

that these differences remained at Wave 3. In addition, significant differences emerged between the two groups in linear growth, with the rate of growth higher in the at-risk than not-at-risk group. In summary, children not at risk had higher levels of performance at Wave 3 but lower linear growth estimates than children at risk.

Also presented in Table 4 for the conditional model are the model random effects for the total sample. When compared with the unconditional model, the estimated variances between subjects for the intercept (.14) and slope (.02) were significant. The random effects (estimated intercept variance between subjects and estimated variance within subjects) in the conditional model can be compared with the random effects in the unconditional model that has no ability group effects. Thus, one way of measuring how much of the variation in word problem solving accuracy was explained by creating ability group comparisons (in the conditional model) was to compute how much of the variance was reduced after comparing the variance estimates of the conditional model with those of the unconditional model. The percentage of reduction in between-subjects variance in the intercept related to the ability group conditional model was 22%: $(.18 - .14)/.18 = .22 \times 100$. We interpret this value by saying that 22% of explainable variation in child intercept values in Wave 3 is a function of the at-risk classification in Wave 1. (Note: the percentage reflected the fraction of variation that was explained. This is not the same as a traditional R^2 statistic.) The explainable variation in slope was 33%: $(.03 - .02)/.03$.

Because of the number of estimates, Table 5 summarizes the fixed effects related to intercept and growth estimates for the ability groups across all measures. As shown in Table 5, the majority of estimates were significantly better than chance. The F ratios are shown comparing the ability groups on estimates of level performance and linear growth. Two important findings emerged. First, the level of performance (intercept values) was significantly lower for children at risk for SMD than children not at risk across all measures. Second, significant differences emerged in growth rates in favor of the not-at-risk group on all measures, except for the reading and phonological knowledge measures. No significant differences emerged between the two groups for growth rates on measures of fluency and STM (phonological loop).

Conditional Model: WM Effects

One critical question that directed this study was whether growth in WM was related to growth in math problem solving accuracy. Table 6 shows the contribution of WM growth to problem solving accuracy. In this model, we tested whether the intercepts and slopes of problem solving were significantly related to the covariates. Covariates in this case were the intercept values and growth estimates for the three WM components. As shown in Table 6, a significant covariate effect was found for level of performance on all three components of WM. The estimated coefficients in Wave 3 for the phonological loop (STM; .14), executive processing (.09), and visual-spatial sketchpad (.10) were significantly related to problem solving. With respect to estimated growth rates, a significant effect occurred for the covariate of the phonological loop (−.05) and executive processing (−.10). The significant parameter estimate related to the phonological loop (−.05) indicated that a child who differed by a score of 1.0 with respect to performance on measures of the phonological loop had a growth rate that differed by −.05. Likewise, a child who differed by a score of 1.0 with respect to the executive system had a growth rate that differed by −.10.

When compared with the unconditional model in Table 3, the conditional model in Table 6 with WM components reduced variance related to between-children intercepts by 61% $(.18 - .07)/.18$ and slopes by 66% $(.03 - .01)/.03$. We determined whether this conditional model provided a good fit to the data by calculating its deviance value (i.e., lack of correspondence between model and data). The log likelihood $-2\ln(L)$ was 1,252.3 for the unconditional model and 954.7 for the conditional growth model. A significant chi-square indicated that the conditional model showed a better fit to the data than the unconditional model, $\chi^2(6) = 297.6, p < .001$.

In summary, ability group effects related to the level of performance in the final wave were higher for children not at risk across all measures when compared with children at risk for SMD. Significant growth differences in favor of the not-at-risk group emerged on all measures, except reading and phonological knowledge. No significant group differences emerged in growth rates on

Table 5
Fixed Effects for the Conditional Model Comparing Intercept and Growth for Children at Risk for Serious Math Difficulties (SMD) and Not at Risk on Measures of Achievement and Cognition

Measure	Group				F
	At risk for SMD		Not at risk for SMD		
	Estimate	SE	Estimate	SE	
Math calculation					
Intercept	1.09***	.07	1.77***	.05	58.10***
Growth	0.58***	.02	0.76***	.02	28.27***
Reading					
Intercept	0.68***	.05	1.00***	.04	34.27***
Growth	0.48**	.02	0.34***	.01	32.25**
Phonological knowledge					
Intercept	0.27**	.05	0.70***	.04	43.48***
Growth	0.28**	.01	0.21**	.01	9.97***
Fluency					
Intercept	0.12**	.04	0.35***	.02	20.72***
Growth	0.13***	.01	0.10***	.01	1.68
Problem solving component					
Intercept	0.54***	.04	0.97***	.03	72.61***
Growth	0.33***	.02	0.46***	.01	22.22***
Speed					
Intercept	-0.69***	.05	-0.86***	.04	7.53**
Growth	-0.50***	.02	-0.21***	.02	85.62***
Inhibition					
Intercept	0.91***	.07	1.35***	.05	23.98***
Growth	0.49***	.02	0.57***	.02	4.91*
Phonological loop					
Intercept	-0.02	.04	0.32***	.03	44.54***
Growth	0.08**	.01	.10***	.01	1.03
Sketchpad					
Intercept	0.15*	.07	0.53***	.05	19.04***
Growth	0.11***	.03	0.25***	.02	11.34***
Executive					
Intercept	0.36***	.05	0.81***	.04	38.78***
Growth	0.20***	.02	0.31***	.02	14.27***

* $p < .05$. ** $p < .01$. *** $p < .001$.

measures of fluency and STM (phonological loop). An advantage in growth was found for the at-risk group on measures of phonological knowledge and reading. However, when the total sample was considered, growth in the phonological loop and the central executive component of WM were related to growth in problem solving accuracy. In addition, performance levels on all three components of WM were significantly related to problem solving accuracy.

Hierarchical Regression

The final analysis determined those cognitive and achievement measures in Wave 1 that best predicted problem solving in Wave 3. These analyses specifically addressed the question as to which components of WM were predictive of problem solving accuracy and whether those predictions were mediated by individual differences in knowledge base, phonological processing, and/or a central executive system. Also included in this part of the analysis for the comparative purposes were the Wave 3 criterion measures of math calculation and knowledge of problem solving components. The estimates for these measures are shown in Table 7.

Prior to the regression analysis, we examined the correlations among the latent measures and the age variable. Correlations between scores for Waves 1 and 3 are shown in Table 8. The majority of

correlations greater than .20 in magnitude were significant at the .001 alpha level. As shown, all Wave 1 measures were significantly related to Wave 3 measures of problem solving accuracy, math calculation, and knowledge of the problem solving components.

An inspection of Table 8 revealed three important findings. (To interpret the results, we considered $r_s > .40$ as substantial correlations.) First, problem solving accuracy in Wave 3 was substantially related to Wave 1 measures of reading, phonological knowledge, fluency, speed, phonological loop, and executive processing. Second, calculation skill and knowledge of problem solving components in Year 3 were strongly correlated with all Wave 1 measures except for inhibition. Finally, high correlations emerged between reading in Wave 1 and measures of math calculation and knowledge of problem solving components in Wave 3.

Predictions of Problem Solving

We investigated whether the relationship between problem solving in Wave 3 and WM components in Wave 1 was maintained when blocks of variables such as knowledge base (e.g., calculation skill), phonological processing (STM, phonological knowledge), and age were entered into the regression analysis. For comparative purposes, we also considered in the analysis Wave 3 criterion

Table 6
Contribution of Working Memory Growth to Problem Solving

Conditional model	Estimate	SE	t
Fixed effects			
Intercept	.47	.02	17.49***
Linear growth	.21	.01	14.51***
Working memory			
Intercept			
STM	.14	.03	3.50**
Executive	.09	.02	3.23***
Sketchpad	.10	.02	4.50***
Growth (linear)			
STM	-.05	.02	-2.08*
Executive	-.10	.01	-5.51***
Sketchpad	.005	.01	0.36
	Variance estimate	SE	z
Random effects (Participants × Teachers)			
Intercept	.07	.01	5.55**
Growth (linear)	.01	.004	3.95**
Residual	.09	.008	13.62**

Note. STM = short-term memory.
* $p < .05$. ** $p < .01$. *** $p < .001$.

measures of math calculation ability and knowledge of problem solving components.

For our first set of analyses, we determined the amount of variance in problem solving performance in Wave 3 that was accounted for by WM components alone in Wave 1. For each model, we entered predictor variables from Wave 1 into the equation simultaneously, so that beta values reflected unique variance (the influence of all other variables was partialled out). For Model 1, we assumed that if scores related to STM (phonological loop) and visual-spatial sketchpad were partialled out in the regression analysis, then the residual variance related to WM could be attributed to the central executive system (e.g., Engle et al., 1999). As shown in Model 1 in Table 9, all components of WM contributed significant variance to the Wave 3 criterion measures. Performance related to the components of WM in Wave 1 contributed to approximately 36% of the variance to problem solving, 37% of the variance to arithmetic calculation, and 36% to problem solving component knowledge in Wave 3. As a follow-up to this model, we considered percentage of variance accounted for in the predictions when only the central executive component of WM was entered into the regression model. The central executive component of WM accounted for 27% of the variance for problem solving accuracy ($R^2 = .27$), $F(1, 291) = 110.06$, $p < .001$, 28% for math calculation ($R^2 = .28$), $F(1, 291) = 112.06$, $p < .001$, and 29% for problem solving knowledge ($R^2 = .29$), $F(1, 291) = 116.31$, $p < .001$. Thus, the contribution of the storage components of WM (phonological loop and visual-spatial sketchpad) in Model 1 increased the percentage accounted for by 9%, 9%, and 7%, respectively, when predicting problem solving accuracy, calculation accuracy, and problem solving knowledge.

In Model 2, we determined whether the contribution of the component scores related to WM in predicting the criterion measures were merely a function of chronological age at Wave 1. As shown in Table 9, all components of WM maintained significant

predictions of the criterion measures when age was partialled out in the analysis. The predictor variables in Model 2 for Wave 1 contributed approximately 38% of variance in word problem solving accuracy, 52% of the variance in math calculation, and 71% of the variance in problem solving component knowledge.

Role of Executive Processes

As shown in Table 9, Model 3 assessed whether the contribution of WM to problem solving remained significant when cognitive

Table 7
Standardized Estimates Used to Create Latent Measures in the Regression Analysis

Domain	Estimate	t
Problem Solving Accuracy (Wave 3)		
Mental Computation: WISC-III	.96	10.14***
Word Problems: Semantic Structure	.25	2.75**
Math (Wave 3)		
WRAT	.92	5.32***
WIAT	.96	4.81***
Computational Fluency	.42	15.58***
Reading (Wave 1)		
WRAT	.93	6.48***
TOWRE: real words	.96	5.94***
Comprehension: WRMT	.94	6.08***
Problem Solving Components (Wave 3)		
a. Question	.78	9.21***
b. Numbers	.83	8.25***
c. Goal	.81	9.05***
d. Operations	.89	9.88***
e. Algorithm	.90	9.06***
Phonological Knowledge: Wave 1		
Pseudowords: TOWRE	.91	9.82***
Elision (Segmentation): CTOPP	.71	7.65***
Naming Speed: Wave 1		
Digit Naming Speed: CTOPP	.95	6.79***
Letter Rapid Naming: CTOPP	.94	6.71***
Phonological Loop (STM): Wave 1		
Digit Forward: WISC-III	.37	12.48***
Pseudoword Span	.45	9.40***
Real Word Span	.72	18.55***
Digit Backward: WISC-III	.42	6.08***
Executive System (working memory): Wave 1		
Update	.58	25.80***
Listening Sentence Span	.50	23.07***
Digit/Sequence Span	.49	24.48***
Semantic Association	.30	14.86***
Visual-Spatial Sketchpad: Wave 1		
Visual Matrix	.94	45.65***
Mapping/Direction	.14	6.29***
Inhibition: Wave 1		
Random Generation Letters	.58	6.59***
Random Generation Numbers	.84	9.66***
Fluency: Wave 1		
Categorical Fluency	.57	5.03***
Letter Fluency	.61	7.15***
Age	1.00	142.22***

Note. Standardized factor loadings represent partial regression coefficients. STM = short-term memory; WISC-III = Wechsler Intelligence Scale for Children—Third Edition; WRAT = Wide Range Achievement Test; WIAT = Wechsler Individual Achievement Test; TOWRE = Test of Word Reading Efficiency; CTOPP = Comprehensive Test of Phonological Processing.

All loadings were statistically significant: ** $p < .01$. *** $p < .001$.

Table 8
Intercorrelations Among Mathematics, Reading, and Cognitive Processing Variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12
Year 1												
1. Age	—											
2. Reading	.62	—										
3. Phonological Knowledge	.48	.89	—									
4. Fluency	.32	.49	.50	—								
5. Inhibition	-.09	-.13	-.11	.01	—							
6. Speed	-.52	-.73	-.67	-.43	.16	—						
7. Sketch Pad	.28	.32	.27	-.20	-.03	-.24	—					
8. Phonological Loop	.24	.47	.45	.32	-.08	-.42	.11	—				
9. Executive	.36	.58	.60	.43	-.16	-.45	.23	.52	—			
Year 3												
10. Problem Solving Accuracy	.37	.59	.55	.40	.01	-.43	.36	.41	.53	—		
11. Calculation	.60	.80	.68	.42	-.03	-.60	.34	.45	.53	.66	—	
12. Problem Knowledge ^a	.78	.79	.64	.44	-.06	-.61	.33	.44	.54	.57	.76	—

^a Knowledge of problem solving components.
rs > .21, *p* < .0001.

variables (fluency, inhibition) assumed to underlie executive processing were entered into the analysis. That is, if the central executive system of WM was primarily related to inhibition and/or fluency, then entering these variables into the regression model should have partialled out the influence of the central executive component of WM. Model 3, when compared with Model 2, increased the percentage of variance accounted for by 2% in predicting word problem solving accuracy, by 6% in predicting calculation, and by 2% in predicting problem solving component knowledge. Wave 1 variables that contributed unique variance to all three-criterion measures were components of WM related to the central executive system. Chronological age contributed unique variance to calculation and problem solving component knowledge but not problem solving accuracy. In general, we did not find support for the notion that measures of fluency, speed, and inhibition eliminated the significant contribution of the central executive system to problem solving

Role of Phonological Processes

Model 4, as shown in Table 9, determined whether the variables related to reading and phonological processing (e.g., reading, phonological knowledge) partialled out the influence of WM in predicting problem solving accuracy, math calculation, and knowledge of problem solving components. When compared with Model 3, Model 4 increased the percentage of variance accounted for by 4% for problem solving accuracy, by 11% for calculation, and by 5% for knowledge of problem solving components. The results showed that when compared with Model 3, the influence of naming speed was partialled out with the entry of reading variables into the equation. The only Wave 1 variable that contributed unique variance to all three-criterion measures was reading. Chronological age contributed unique variance to calculation and problem solving component knowledge. As found in Model 3, the central executive component of WM contributed unique variance to word problem solving accuracy and knowledge of problem solving components. In general, the results suggest that when compared with the previous models, the significant effects of the central

executive system of WM were eliminated only on the measure of calculation when measures of reading were entered into the regression analysis.

Role of Knowledge Base

Admittedly, we were surprised that measures of the central executive component of WM were not eliminated in predicting later performance of problem solving accuracy when measures of reading and executive processing activities (e.g., inhibition) were entered into the analysis. We reasoned that, perhaps, domain-specific knowledge in math calculation and general reasoning ability may account for the robust findings related to the WM measures. Thus, in a final model (Model 5) we entered Wave 3 scores related to math calculation ability and knowledge of problem solving components, and Wave 1 scores for fluid intelligence (Raven Colored Progressive Matrices Test) into the regression model. This model directly tested whether individual differences in children’s knowledge base underlie the relationship between WM and problem solving. The results are shown in Table 10. The model determined whether domain-specific knowledge (math) and reasoning ability (fluid intelligence) would partial out the effects of WM in predicting problem solving accuracy. When compared with Model 4, the final model increased the percentage of variance accounted for by 5% when predicting problem solving accuracy. Predictor variables in Wave 3 that contributed significant variance to problem solving accuracy were math calculation and problem solving knowledge, and predictor variables in Wave 1 that contributed significant variance were fluid intelligence, visual–spatial sketchpad, and the central executive component of WM. Interestingly, when compared with Model 4, Model 5 completely eliminated the significant influence of reading on problem solving accuracy.

In summary, there were two important findings related to the hierarchical regression analysis. First, entering various cognitive variables attributed to executive processing (speed, fluency, and inhibition) as well as reading skill into the regression model did not eliminate the contribution of executive processing to predicting

Table 9

Hierarchical Analysis Predicting Word Problem Solving, Math Calculation, and Problem Solving Component Knowledge in Wave 3 From Cognitive and Achievement Variables in Wave 1

Regression modeling	Word problem solving				Calculation				Problem solving components			
	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>B</i>	<i>SE</i>	β	<i>t</i>
Model 1												
Sketchpad	.26	.04	.26	5.47***	.23	.04	.23	4.77***	.22	.04	.22	4.49***
Phonological loop	.20	.06	.17	3.02***	.37	.06	.22	3.92***	.25	.06	.20	3.55***
Executive	.48	.07	.37	6.63***	.47	.07	.36	6.39***	.49	.07	.38	6.70***
Model 2												
Age	.14	.04	.14	2.75***	.43	.04	.44	9.89***	.63	.03	.65	18.69***
Sketchpad	.23	.04	.23	4.80**	.14	.04	.13	3.22***	.08	.03	.08	2.44*
Phonological loop	.19	.06	.15	2.81**	.22	.05	.17	3.67***	.17	.04	.14	3.36**
Executive	.44	.07	.33	5.90***	.32	.06	.25	4.93**	.28	.04	.22	5.41***
Model 3												
Fluency	.14	.07	.10	1.94	.10	.06	.08	1.75	.11	.04	.08	2.45*
Speed	-.10	.06	-.11	-1.83	-.23	.05	-.23	-4.34***	-.14	.04	-.14	-3.50**
Inhibition	.08	.08	.06	1.28	-.02	.07	-.01	-0.35	-.01	.05	-.01	-0.38
Age	.08	.05	.08	1.54	.34	.04	.35	7.75***	.57	.03	.59	16.13***
Sketchpad	.21	.04	.22	4.42**	.11	.04	.11	2.71**	.05	.03	.05	1.67
Phonological loop	.16	.06	.12	2.21***	.14	.05	.15	2.35**	.11	.04	.09	2.38**
Executive	.37	.07	.29	4.78***	.23	.06	.17	3.32**	.20	.04	.16	3.84**
Model 4												
Reading	.34	.12	.34	2.75**	.68	.09	.68	7.29***	.52	.07	.53	6.85***
Phonological knowledge	-.01	.11	-.01	-0.01	-.09	.08	-.09	-1.15	-.18	.06	-.17	-2.56*
Fluency	.07	.07	.06	1.21	.02	.05	.01	0.38	.07	.04	.05	1.60
Speed	.02	.06	.02	0.36	.01	.05	.01	0.24	.002	.04	.002	0.05
Inhibition	.07	.08	.05	1.08	-.04	.06	-.03	-0.96	-.03	.05	-.02	-0.88
Age	-.01	.05	-.01	-0.17	.17	.04	.17	3.81**	.45	.03	.46	12.40***
Sketchpad	.20	.04	.20	4.16***	.08	.03	.08	2.25*	.03	.03	.03	1.13
Phonological loop	.13	.06	.10	1.84	.09	.05	.07	1.68	.09	.04	.06	1.75
Executive	.27	.08	.21	3.40**	.06	.05	.04	1.00	.11	.05	.09	2.30*

Note: Model 1, $F(3, 289) = 54.82, p < .001, R^2 = .36$; $F(3, 289) = 55.51, p < .001, R^2 = .37$; $F(3, 289) = 54.20, p < .001, R^2 = .36$. Model 2, $F(4, 288) = 43.94, p < .001, R^2 = .38$; $F(4, 288) = 80.05, p < .001, R^2 = .52$; $F(4, 288) = 176.92, p < .001, R^2 = .71$. Model 3, $F(7, 284) = 27.30, p < .001, R^2 = .40$; $F(7, 284) = 52.62, p < .001, R^2 = .57$; $F(7, 284) = 110.75, p < .001, R^2 = .73$; Model 4, $F(9, 282) = 24.23, p < .001, R^2 = .44$; $F(9, 282) = 65.72, p < .001, R^2 = .68$; $F(9, 282) = 111.38, p < .001, R^2 = .78$.

* $p < .05$. ** $p < .01$. *** $p < .001$.

problem solving accuracy or performance on the knowledge of problem solving components task. Second, performance on different components of WM predicted different criterion measures. The executive system and visual-spatial sketchpad were important predictors of problem solving accuracy, whereas performance related to the visual-spatial sketchpad predicted calculation, and the executive system predicted performance on the knowledge of problem solving components task.

Discussion

The purpose of this study was to identify the cognitive processes in WM that predict mathematical problem solving performance in elementary school children who are at risk and not at risk for serious math difficulties. We determined whether WM was a valid construct in the predictions of problem solving and whether measures of WM predicted problem solving above and beyond children's performance on measures related to reading, phonological processing, computational knowledge, and domain-specific knowledge. We also determined whether growth in WM was related to growth in problem solving accuracy. Overall, our data provide substantial support for the notion that the capacity to store and process material in WM significantly constrains a child's

ability to problem solve during the elementary school years. Further, children identified at risk for problem solving difficulties in Wave 1 maintained their risk status on measures of problem solving and WM 2 years later. In the discussion that follows we review our findings related to WM and problem solving. We then discuss the implications of our findings for future research and practice.

Testing of Three Models

We tested three models of WM and its influence on growth in problem solving. One model tests whether phonological processes (e.g., STM-phonological loop, phonological knowledge) play a major role in predicting performance in problem solving and whether the phonological system mediates the influence of executive processing (WM) on problem solving. Phonological processes in this study were related to latent measures of STM and phonological knowledge (measures of elision, pseudoword reading). The model follows logically from the reading literature that links phonological skills to new word learning (e.g., Baddeley et al., 1998), comprehension (Perfetti, 1985), and mental calculation (e.g., Logie, Gilhooly, & Wynn, 1994). The model assumes that poor problem solvers have deficits in the processing and storage of

Table 10
Predictions of Year 3 Problem Solving Accuracy Based on Wave 3 Math Calculation, Problem Solving Knowledge, and Wave 1 Fluid Intelligence, Reading, and Cognitive Variables

Model 5 predictor	<i>B</i>	<i>SE</i>	β	<i>t</i>
Wave 3				
Problem Solving Knowledge	.25	.11	.12	2.13*
Calculation	.30	.08	.27	3.42**
Wave 1				
Fluid Intelligence (Raven)	.13	.04	.16	2.85**
Reading	.12	.12	.12	1.00
Phonological Knowledge	-.01	.10	.10	-0.09
Fluency	.02	.07	.007	0.33
Speed	-.004	.06	-.004	-0.06
Inhibition	.09	.06	.07	1.60
Age	-.15	.06	-.16	-2.39*
Sketchpad	.15	.04	.14	3.23***
Phonological Loop	.12	.06	.09	1.85
Executive	.19	.08	.15	2.34*

Note. $F(12, 279) = 22.52, p < .001, R^2 = .49$.

* $p < .05$. ** $p < .01$.

phonological information that creates a bottleneck in the flow of information to higher levels of processing.

The second model suggests that domain-specific knowledge in LTM mediates individual differences in WM and problem solving. Measures of LTM in this study were related to calculation ability, reading, and knowledge of problem solving components. It has recently been argued that some of the functions of the central executive system are to access information from LTM (e.g., Baddeley & Logie, 1999), and, therefore, this model suggests that controlling for the activation of LTM (e.g., arithmetic calculation, knowledge of algorithms) would partial out the influence of WM on problem solving.

The final model suggests that problem solving performance relates to executive processing, independent of the influence of the phonological system and LTM. This assumption follows logically from the problem solving literature suggesting that abstract thinking, such as comprehension and reasoning, requires the coordination of several basic processes (e.g., Engle et al., 1999; Just, Carpenter, & Keller, 1996; Kyllonen & Christal, 1990). We assumed that measures of executive processing in this study were related to latent measures of WM, and measures assumed to reflect some of the activities of the central executive system were inhibition (random generation of letters and numbers) and activation of LTM (composite measures of reading, arithmetic calculation, knowledge of problem solving components). We also assumed based on the work of Engle et al. (1999) and others (e.g., Cowan, 1995) that after storage processes (phonological loop and visual-spatial) were partialled from the analysis that the residual variance related to WM captured a key process of the central executive system referred to as *controlled attention*. The results yield two clear findings in support of the third model. However, when interpreting these findings the reader needs to keep in mind that other processes not reflected in our latent measures may also play a role.

First, WM contributes unique variance to problem solving beyond what phonological processes (e.g., phonological knowledge), reading skill, inhibition, and processing speed contribute. The

results show that WM performance in Wave 1 contributed approximately 36% of the variance to problem solving accuracy in Wave 3 when entered by itself in the regression analysis. Thus, there is clear evidence that multiple systems of WM contribute important variance to problem solving performance beyond processes related to speed, phonological knowledge, and reading skill.

Second, WM performance predicted problem solving accuracy when the hierarchical regression analysis included measures of LTM. Measures of LTM in this study comprised the tasks related to reading, calculation, and knowledge of problem solving components. It has been argued that some of the functions of the central executive system include accessing information from LTM (e.g., Baddeley & Logie, 1999; Conway & Engle, 1994). Our results suggest, however, that although WM tasks draw information from LTM (e.g., knowledge of components related to problem solving), it may be the controlled attention (monitoring of attention) component of WM that plays a more important role in mathematical problem solving growth. The results clearly show in the final model (see Table 10, Model 5) that neither knowledge of problem solving components nor calculation ability partialled out the significant influence of WM in predicting mathematical problem solving accuracy. We will now address the three specific questions that directed our study.

1. Do children identified at risk or not at risk in Wave 1 vary on measures of growth in problem solving and WM across the testing waves?

To answer this question, we estimated intercept values in Wave 3 and the rate of growth across testing waves on measures of problem solving and WM. The growth curve analysis showed that ability group differences in problem solving emerged for the estimated intercept values at Wave 3. Thus, although ability group differences would be expected at Wave 1, the results showed that these differences were sustained at Wave 3. Although we were uncertain whether improvement in reading would mitigate any word problem solving differences in subsequent years, children at risk for SMD in Wave 1 were also at risk in Wave 3. In addition, significant differences emerged between the two groups in linear growth, with the at-risk group showing a lower growth rate when compared with the not-at-risk group. We also found that the intercept and linear growth parameters were significantly different between children at risk for SMD and not at risk on measures of math calculation and the executive component of WM. Children at risk for SMD had lower intercepts and smaller growth rates than children not at risk. Interestingly, no differences were found between the ability groups on growth estimates for performance on measures of the phonological loop and fluency.

We also compared the two groups across a broad array of achievement and cognitive measures. The general pattern in the repeated measures ANCOVA for the intact sample across all three waves was that the comparisons between at-risk and not-at-risk students were significant at each wave and for each academic and cognitive domain. However, greater differences emerged between ability groups for some domains than others. A comparison of groups across achievement and cognitive domains (collapsed across time) indicated that greater group differences in favor of children not at risk emerged on measures of problem solving accuracy, math calculation, reading, and speed when compared

with the cognitive variables (fluency, inhibition, and memory components; e.g., phonological loop) and knowledge of problem solving components. The general trend was that children not at risk scored higher than those at risk for SMD, and higher difference scores emerged for academic domains (e.g., reading, math calculation) than cognitive domains.

2. Is growth in WM related to growth in word problem solving accuracy?

Our multilevel analysis showed a positive effect for growth in WM and problem solving accuracy. The general findings of the growth analysis were that ability group effects related to intercepts and growth parameters emerged for measures of problem solving. Children at risk for SMD showed lower levels of performance and less linear growth rate than children not at risk. When the total sample was analyzed, however, another important finding was that all intercept values related to components of WM (phonological loop, visual-spatial sketchpad, and central executive) were significantly related to problem solving. A significant relationship was found for the level of performance on all three WM components to problem solving accuracy. However, only the linear growth in the phonological loop and executive processing was related to word problem solving.

No doubt, it could be argued that growth rates in problem solving are merely a function of classroom instruction. We addressed this issue by comparing an unconditional and conditional model that included growth of the students nested within classroom with and without the inclusion of WM. The conditional model that included WM reduced the influence of classroom (teacher) variance on between-children intercepts by 61% and between-children growth by 66%. Thus, we found that a large percentage of the variance related to participant performance nested within classroom instruction on problem solving could be accounted for by individual differences in WM.

3. Which components of WM were predictive of problem solving accuracy and are those predictions mediated by individual differences in knowledge base, phonological processing, and/or the central executive system?

We tested this question by utilizing a hierarchical regression analysis, in which variables related to WM, cognitive processing (speed, inhibition, fluency), and achievement (e.g., reading, phonological knowledge) in Wave 1 were systematically entered into the regression equation to predict problem solving in Wave 3. There were two important findings.

First, the results show that the executive component of WM contributes unique variance to problem solving accuracy beyond what phonological processes (e.g., STM, phonological knowledge) contribute. The results show that the executive component of WM performance in Year 1 contributes approximately 27% of the variance to problem solving accuracy in Wave 3 when entered by itself in the regression analysis.

Second, the executive component of WM performance contributes unique variance to problem solving accuracy even when the hierarchical regression includes measures of LTM. Our results suggest, however, that although WM tasks draw information from LTM (e.g., knowledge of components related to problem solving),

it may be the monitoring of the components in WM system (e.g., manipulation of information) and storage of visual-spatial information that play a more important role in predicting later performance on mathematical problem solving measures. However, our results must be clarified because we have merely partialled out the influence of the amount of knowledge available. A distinction must be drawn between the amount of WM resources available to produce activation and the amount of knowledge that can be activated. Thus, in our regression analysis we assessed the quality of information available, not the capacity to activate LTM knowledge. The results merely show that both knowledge of problem solving components and calculation ability in Wave 3 did not partial out the significant influence of WM in Wave 1 in predicting word problem solving accuracy in Wave 3.

As a parallel to predicting word problem solving accuracy, we also considered predictions related to mathematical computation. In terms of math calculation in Wave 3, the best Wave 1 predictors were reading, chronological age, and the visual-spatial storage (sketchpad) component of WM. Thus, in contrast to the important role that the executive component of WM plays in problem solving accuracy, it appears that performance in the visual-spatial sketchpad also predicts math calculation. The visual-spatial sketchpad is specialized for the processing and storage of visual material, spatial material, or both and for linguistic information that can be recoded into visual forms (see Baddeley, 1986, for a review). Our findings are consistent with studies (Gathercole & Pickering, 2000a, 2000b) finding that visual-spatial WM abilities, as well as measures of executive processing, were associated with attainment levels on a national curriculum for children ages 6 to 7 years. Children who showed marked deficits in curriculum attainment also showed marked deficits in visual-spatial WM. Thus, there is a strong relationship between measures of the visual-spatial sketchpad and academic performance in the younger grades.

Comparisons With Related Studies

Not all studies have found a significant relationship between WM and problem solving. For example, Fuchs et al. (2006) studied the cognitive correlates of arithmetic computation and arithmetic word problem ability of third graders. Predictor variables were language, nonverbal problem solving, concept formation, processing speed, long-term memory, WM, phonological decoding, and sight word reading efficiency. Word problem accuracy was best predicted by concept formation and sight word efficiency. The only significant cognitive predictors of arithmetic were phonological decoding and processing speed. Likewise, earlier work by Fuchs et al. (2005) showed that phonological processing was a unique determinant of the development of arithmetic skills in first graders. The study by Hecht, Close, and Santisi (2003) also demonstrated that phonological processing almost completely accounted for the association between reading and computational skill in older children. In contrast, Lee, Swee-Fong, Ee-Lyn, and Zee-Ying (2004) assessed the performance of 10-year-olds on measures of word problems, WM, intelligence, and reading ability. They found that children who had greater storage and greater central executive system capacities related to WM measures were better able to solve mathematical problems. At the same time they found that children with higher IQs, reading proficiencies, and vocabularies, performed better on mathematical problems. Further,

although WM was found to be related to problem solving, WM did not contribute important variance to problem solving accuracy after measures of reading were entered into the regression analysis. This finding replicated an earlier finding from the study by Swanson et al. (1993), who found that reading comprehension performance superseded any unique contributions made by WM measures to problem solving (also see Kail & Hall, 1999, for a similar finding). These findings are in contrast to other researchers (e.g., De Beni, Palladino, Pazzaglia, & Cornoldi, 1998; Passolunghi, Cornoldi, & De Liberto, 2001), who have shown that when controlling for vocabulary and reading, poor problem solvers exhibit lower WM scores (as indexed by their poorer inhibitory abilities) than do good problem solvers.

Thus, there is some confusion in the literature about the role of WM in problem solving. To examine these issues, we give attention to the recent study by Fuchs et al. (2006). As stated, Fuchs et al. (2006) failed to find that WM predicted word problems accuracy, and this finding varies considerably from our findings. There are three major distinctions between the Fuchs et al. (2006) study and our study. First, there were variations in WM measures. Fuchs et al. (2006) primarily relied on two WM measures. One was the backward span that some consider a STM rather than WM measure (Colom, Abad, et al., 2005; Colom, Flores-Mendoza, et al., 2005; also see Footnote 1). In contrast, our study used a comprehensive battery of WM measures. Second, participants in the Fuchs et al. (2006) study were primarily sampled from Title 1 schools and identified on the basis of the performance on a nonstandardized test of computational fluency. Further, approximately 60% of their sample received subsidized lunch. In contrast, the present study sample of children comes from mostly middle-class or upper middle-class homes who were selected on a standardized measure of problem solving. In addition, it is important to note that the at-risk sample in the present study included students who were within the average range in arithmetic and reading skills. In contrast to the Fuchs et al. (2006) study that defined risk as the ability to calculate, our study defined risk as the ability to problem solve and name numbers quickly. Finally, as indicated by one reviewer, our results when compared with the work of Fuchs et al. (2006) were primarily due to methodology differences. For example, variations emerged in the criterion measure. In the Fuchs et al. (2006) study, word problems were presented as text and students had to process text. Children were allowed 3 s to respond after hearing the problem but could reread the passages if necessary. This activity would certainly lessen the demands on WM capacity because children can reread the sentence. In contrast, we read the problem to the students and students were able to focus on the problem being read while maintaining information in WM. Therefore, the reviewer was not surprised that reading played a minimal role in our study and was important in the Fuchs et al. (2006) study. Further, as indicated by the reviewer, the presentation of word problems in text in the Fuchs et al. (2006) study may be more ecologically valid than ours. We would argue that it is equally valid to problem solve without recourse to text—thinking on your feet as it were. Everyday problems are not always presented in text and it may be ecologically valid to also teach problem solving without having to rely on printed text. As indicated by an anonymous reviewer of this article, we could argue that we have a purer

measure of word problem solving because reading was not required.

On the issues of reading, some studies have argued that reading proficiency mediates the relationship between WM and problem solving. For example, Lee et al. (2004) showed that literacy measures provided greater contribution than measures related to the central executive system to word problem solving ability in children. Although WM was significantly related to word problem solving ability, its variance was significantly reduced when the influence of reading was partialled out. Further, they found that the storage component of WM, primarily the phonological loop and the visual-spatial sketchpad, did not contribute directly to mathematical performance.

Similar to Lee et al. (2004), as well as Fuchs et al.'s (2006) study, we found that reading was an important predictor of calculation. However, the contribution of WM to problem solving in our study was larger than that of reading. No doubt, the weak association of reading to problem solving in Model 5 (see Table 9) could be explained by the fact that all problems used in the current study were read to the students; consequently, reading was weakened as a possible mediating variable. We argue, however, that if children are unable to understand and decode a question, further processing is unlikely to lead to a correct solution. Therefore, reading abilities may account for the similarities and differences in the findings across studies.

In addition, there are other studies that clearly show that reading or reading-related processes do not directly mediate the influence of WM on problem solving. For example, Swanson and Sachse-Lee (2001) found with 12-year-old math-disabled and chronologically age-matched peers that phonological processing, WM (executive component), and visual WM contributed unique variance to problem solving accuracy. Thus, they did not find support that reading ability or literacy processes mediated the role of WM in the problem solving accuracy. In a follow-up study, Swanson (2004) compared two age groups (7 and 11) on WM and problem solving measures. This study found that regardless of age, WM predicted problem solving accuracy in word problems independent of measures of problem representation, knowledge of operations and algorithms, phonological processing, fluid intelligence, and achievement in reading and math. Further, the results suggest that a central executive system underlies age-related improvements in word problem solving accuracy.

In general, we argue that the correct calculation of numbers presented on paper requires some minimum threshold of reading ability. A lack of control for this confound may lead to attributing difficulties in math to the same cognitive processes (phonological process) as reading. In addition, children who read at a certain threshold (e.g., above the 25th percentile in word identification) may fail to be adequately diagnosed as having potential difficulties in mathematical problem solving. In order to circumvent these problems, we feel it is necessary to identify processes in children at risk for SMD that are distinct from the processes related to reading. Our results clearly show that children at risk for SMD on a nonreading measure but whose performance is in the average range on measures of fluid intelligence, reading skill, and math calculation skill are deficient on measures of WM when compared with children not at risk for problem solving.

Implications

In general, our findings show that WM performance predicts later performance in problem solving accuracy, mathematical calculation, and knowledge of problem solving components. What are the implications of our findings to education of children? We believe the results have three implications.

1. Our classification criteria have predictive validity. Children identified at risk for SMD in Wave 1 were deficient in problem solving in Wave 3. Children identified at risk for SMD in Wave 1 did not catch up to children not at risk in Wave 3 in problem solving, math, and WM. Further, as shown in Figure 1, several other areas were found deficient in children identified at risk for serious mathematical problem solving difficulties (e.g., reading, random generation). Thus, our classification based on digit naming speed and the Arithmetic subtest of the WISC-III (which includes word problems) was a valid discriminator in predicting subsequent growth in Year 3.

Of course, one may criticize our classification criteria both in terms of the tasks we used, especially the Arithmetic subset of the WISC-III, and the cutoff criteria as well. Let us consider three arguments related to the Arithmetic subtest. One argument is that risk status should be based on calculation skill and not problem solving ability. However, we argue that we were able to identify a substantial number of children who were in the average range on reading and calculation measures but had clear difficulties listening to a problem with a verbal text and coming up with a correct solution. We are very much aware of the work by others (e.g., Geary, 2003), indicating that math disability should focus on calculation only. However, selection criterion for risk has varied across studies from the 25th percentile to the 45th percentile (see Swanson & Jerman, 2006, for a review). We think one of the reasons for this variation in cutoff scores is that standardized calculation scores have an extremely restrictive range of items for the younger grades and/or items do not match demands of the classroom. That is, popular standardized measures of calculation place children higher than actual classroom performance (e.g., see Fuchs et al., 2004, p. 496, for discussion of this issue). However, when problem solving difficulties were considered, we were able to sample enough children below a conservative cutoff score of the 25th percentile.

Another argument is that the Arithmetic subtest from the WISC-III was not intended as a measure of math. Although Sattler (2001) and others suggested that the Arithmetic subtest from the WISC-III was not intended to be a math measure, it is clearly correlated with math achievement, as well as vocabulary and other related skills. Regardless of the intent of the measure, the arithmetic subtest requires children to answer simple to complex problems involving arithmetic concepts and numerical reasoning. The first couple of problems require the direct counting of discrete objects. Problems 3, 4, and 5 require subtraction using objects and stimuli. The remaining problems require addition, subtraction, multiplication, and division.

Another argument is that the Arithmetic subtest itself is a WM measure. No doubt this task taps WM as well as prior learning, but it also correlates more highly with information and similarities than the other subtests (e.g., Digit Span). In fact, it has a *g* loading the same as Vocabulary, Information, and Similarities (see Keith et al., 2006, p. 122), whereas Digit Span is substantially lower.

Although we recognize that the mental computation required for this task does place constraints on WM capacity, the goal in our study was to find out which components of WM are the most constrained for children who have difficulties related to mental calculation.

In summary, our finding is important because children who were at risk for problem solving were generally in the average range on reading and math calculation measures but sustained difficulties on a number of measures across all testing waves.

2. A major cognitive component that underlies risk status in elementary school children with average intelligence is the executive component of WM. In contrast to the work on reading, it does not appear to be the case in this study that deficits in WM were merely a manifestation of deficits in the phonological system (also see Swanson & Ashbaker, 2000, for a similar finding). We found in our regression modeling that WM predicted math problem solving even when the influence of STM was partialled out.

The above finding is consistent with other studies showing that WM is an important predictor of various cognitive operations even when the influence of STM is partialled out in the analysis (e.g., Engle et al., 1999). Although we assumed that the residual variance related to WM was related to some aspect of the central executive system, most notably controlled attention, an important point raised by one anonymous reviewer of this article is that we have no direct measure indicating that the residual variance related to WM is controlled attention. Further, the reviewer also indicated that our findings are merely an artifact of *g*. For example, children at risk for SMD had lower Raven scores than those not at risk. However, it is important to note that fluid intelligence was partialled out in the analysis in Model 5. Even under these conditions, WM still contributed significant variance to word problem solving. However, the reviewer is correct that one cannot assume that the residual variance related to WM is controlled attention. Let us consider this point in more detail.

Our results extend Engle et al.'s (1999) findings to children that when controlling for the correlations between WM and STM, the residual variance for the WM factor predicts problem solving. The question emerges, however, as to whether the residual variance attributed to WM and problem solving reflects controlled attention (e.g., Engle et al., 1999), a domain general attentional resource involved in the activation of information from LTM (e.g., Cantor & Engle, 1993), a general monitoring system that coordinates the flow of information but draws from specialized storage systems (e.g., Baddeley & Logie, 1999), or a limited-capacity resource that supports both processing and storage in a domain-specific system (e.g., Just & Carpenter, 1992).

To answer this question, let us first consider Model 3 in Table 9. We assumed that some of the activities related to executive processing component of the WM measure (i.e., controlled attention) were related to the latent measures of fluency, random generation, and speed of processing. However, we found that entry of these measures in the regression models did not eliminate the significant contribution of WM to problem solving accuracy or calculation. What these findings suggest to us is that attentional processes dissimilar from those captured by measures of fluency, speed, and random generation mediate the residual variance related to WM in predicting problem solving. Further, because our measures were not pure measures of inhibition, we may not have adequately tested the inhibition model. Interestingly, in Model 5

(see Table 9), we found that both the executive and visual-spatial component of WM significantly predicted math problem solving when measures of literacy (reading and phonological knowledge) were entered into the regression model. Thus, weak support was found for the phonological model, because the entry of those variables into the regression analysis failed to partial out the influence of WM in predictions of problem solving. Thus, verbal forms of processing efficiency (naming speed) and content (phonological knowledge) did not adequately account for the influence of WM. On the basis of these results, we argue that if WM is made up of controlled attention and storage, then when the influence of storage is partialled out in the analysis, what is left in terms of residual variance is some form of controlled attention. However, given that the random generation, fluency tasks, and measures of LTM did not eliminate the contribution of this residual variance to word problem solving accuracy, we tentatively concluded that some yet-to-be-specified aspects of controlled attentional processing play an important predictive role in elementary school children's problem solving.

3. Limitations in WM capacity can be compensated for by improved fluency and proficiency on academic tasks (i.e., knowledge of problem solving components, reading comprehension, and math computation). Although the influence of individual differences in WM on problem solving performance is robust, this does not mean that its influence cannot be compensated for. Increased performance on measures related to math calculation and knowledge of algorithms reduced the influence of individual differences in WM on problem solving. What remains to be examined is the influence of instruction and age-related development on these processes. From a practical standpoint, the finding that WM tests do mediate problem solving outcomes should be considered when we look at high stakes examinations. Although our research suggests that literacy (reading) is important in terms of augmenting (predicting) word problem solving ability (see Model 4, Table 9), this is not the only variable that needs to be considered when compensating for demands on WM. The demands on reading are not greater than the demands on WM. In fact, the actual percentage of variance contributed to word problem solving ability is higher for the WM measures than it is for the literacy measures. The results do suggest that mathematical calculation skills are important in the study; so clearly there must be a threshold that would be extremely important. It appears that the children in this study had mastered or at least performed above the threshold and, therefore, were able to rely on other resources.

Conclusion

In summary, growth in WM predicts growth on problem solving measures. We believe these results extend those studies on individual differences that suggest that a WM system plays a critical role in integrating information during problem solving. These models explicitly posit a dual role of WM: (a) It holds recently processed information to make connections to the latest input, and (b) it maintains information for the construction of an overall solution to problems. In terms of individual differences, children who have a large WM capacity for language can carry out the execution of various fundamental problem solving processes (such as problem representation, problem execution, etc.) with less demands on a limited resource pool than children with a smaller WM

capacity. As a result, children with a larger WM capacity would have more resources available for storage while representing the problem. On the other hand, children with a smaller WM capacity might have fewer resources available for the maintenance of information during problem solving. Further, this relationship holds (at least for children) even when the influence of phonological loop (naming speed and STM) is partialled from the analysis. Yet, WM is not the exclusive contributor to variance in problem solving ability. This study also supports previous research about the importance of reading skill in solution accuracy (see Model 4). Moreover, our findings are consistent with models of high order processing, which suggest that WM resources activate relevant knowledge from LTM but also suggest that a subsystem that controls and regulates the cognitive system plays a major role. What is unclear from previous studies is whether growth in the WM system is related to growth in problem solving. This study has provided clear evidence suggesting that growth in WM is related to growth in problem solving. Thus, we think one of the core problems children face in solving mathematical word problems relates to growth in operations ascribed to WM.

References

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*, 567-589.
- Anderson, E. R. (1993). Analyzing change in short-term longitudinal research using cohort-sequential designs. *Journal of Consulting and Clinical Psychology, 61*, 929-940.
- Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30*, 221-256.
- Baddeley, A. D. (1986). *Working memory*. London: Oxford University Press.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417-422.
- Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158-173.
- Baddeley, A. D., & Logie, R. H. (1999). The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). Cambridge, England: Cambridge University Press.
- Barrouillet, P., & Lépine, R. (2005). Working memory and children's use of retrieval to solve addition problems. *Journal of Experimental Child Psychology, 91*(3), 183-204.
- Bentler, P. M., (2005). *EQS 6 Structural equations program manual* [Computer software manual]. Encino, CA: Multivariate Software.
- Bentler, P. M., & Wu, E. J. C. (1995). *Structural equations program manual* [Computer software manual]. Encino, CA: Multivariate Software.
- Bryk, A., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.
- Bull, R., Johnston, R. S., & Roy, J. A. (1999). Exploring the roles of the visual-spatial sketchpad and central executive in children's arithmetical skills: Views from cognition and developmental neuropsychology. *Developmental Neuropsychology, 15*, 421-442.
- Byrne, B. M. (2006). *Structural equation modeling with EQS*. Mahwah, NJ: Erlbaum.
- Cantor, J., & Engle, R. W. (1993). Working-memory capacity is long-term memory activation: An individual-differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1101-1114.

- Colom, R., Abad, F., Rebollo, I., & Shih, P. C. (2005). Memory span and general intelligence: A latent-variable approach. *Intelligence, 32*, 623–642.
- Colom, R., Flores-Mendoza, C., Quiroga, M. Á., & Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality & Individual Differences, 39*, 1005–1014.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163–183.
- Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123*, 354–373.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford, England: Oxford University Press.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.
- De Beni, R., Palladino, P., Pazzaglia, F., & Cornoldi, C. (1998). Increases in intrusion errors and working memory deficit of poor comprehenders. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 51(A)*, 305–320.
- Dempster, F. (1985). Short-term memory development in childhood and adolescence. In C. Brainerd & M. Pressley (Eds.), *Basic processes in memory* (pp. 209–248). New York: Springer-Verlag.
- Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 972–992.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309–331.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245.
- Fayol, M., Abdi, H., & Gombert, J. (1987). Arithmetic problem formulation and working memory load. *Cognition and Instruction, 4*, 187–202.
- Fletcher, J. M., Epsy, K. A., Francis, P. J., Davidson, K. C., Rourke, B. P., & Shaywitz, S. E. (1989). Comparison of cutoff and regression-based definitions of reading disabilities. *Journal of Learning Disabilities, 22*, 334–338.
- 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*, 493–513.
- Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*, 29–43.
- Fuchs, L. S., Fuchs, D., Eaton, S., Hamlett, C. L., & Karns, K. (2000). Supplementing teacher judgments of mathematics test accommodations, with objective data sources. *School Psychology Review, 20*, 65–85.
- Fuchs, L. S., Fuchs, D., Prentice, K., Hamlett, C. L., Finelli, R., & Courey, S. J. (2004). Enhancing mathematical problem solving among third-grade students with schema-based instruction. *Journal of Educational Psychology, 96*, 635–647.
- Furst, A., & Hitch, G. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory & Cognition, 28*, 774–782.
- Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry, 39*, 3–27.
- Gathercole, S. E., & Pickering, S. J. (2000a). Assessment of working memory in six- and seven-year-old children. *Journal of Educational Psychology, 92*, 377–390.
- Gathercole, S. E., & Pickering, S. J. (2000b). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology, 70*, 177–194.
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177–190.
- Geary, D. C. (2003). Math disabilities. In H. L. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities*. (pp. 199–212). New York: Guilford Press.
- Geary, D. C., Brown, S. C., & Samaranayake, V. A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology, 27*, 787–797.
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology, 88*, 121–151.
- Harrison, J. E., Buxton, P., Husain, M., & Wise, R. (2000). Short test of semantic and phonological fluency: Normal performance, validity, and test-retest reliability. *British Journal of Clinical Psychology, 39*, 181–191.
- Hecht, S. A., Close, L., & Santisi, M. (2003). Sources of individual differences in fraction skills. *Journal of Experimental Child Psychology, 86*, 277–302.
- Hegarty, M., Mayer, R., & Monk, C. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87*, 18–32.
- Heitz, R. P., Unsworth, N., & Engle, R. W. (2005). Working memory capacity, attention control, and fluid intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 61–78). Thousand Oaks, CA: Sage.
- Henry, L. A., & Millar, S. (1993). Why does memory span improve with age? A review of the evidence for two current hypotheses. *European Journal of Cognitive Psychology, 5*, 241–287.
- Hitch, G. J., & McAuley, E. (1991). Working memory in children with specific arithmetical learning difficulties. *British Journal of Psychology, 82*, 375–386.
- Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood* [Computer software manual]. Mooresville, IN: Scientific Software.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149.
- Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review, 103*, 773–780.
- Kail, R., & Hall, L. K. (1999). Sources of developmental change in children's word-problem performance. *Journal of Educational Psychology, 91*, 660–668.
- Keith, T., Fine, J. G., Taub, G., Reynolds, M., & Kranzler, J. (2006). High order, multisampling, confirmatory factor analysis of the Wechsler Intelligence Scale For Children—Fourth Edition: What does it measure? *School Psychology Review, 35*, 108–127.
- Keith, T., & Wittam, E. (1997). Hierarchical and cross-age confirmatory factor analysis of the WISC-III: What does it measure? *School Psychology Review, 12*, 89–107.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving arithmetic word problems. *Psychological Review, 92*, 109–129.
- Klauer, K. J., Willmes, K., & Phye, G. (2002). Inductive reasoning: Does it transfer to fluid intelligence? *Contemporary Educational Psychology, 27*, 1–25.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence, 14*, 389–433.
- Lee, K., Swee-Fong, N., Ee-Lynn, N., & Zee-Ying, L. (2004). Working

- memory and literacy as predictors of performance on algebraic word problems. *Journal of Experimental Child Psychology*, 89, 140–158.
- Logie, R. H., Gilhooly, K. J., & Wynn, M. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition*, 22, 395–410.
- Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problem solving. In R. J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp. 29–54). Mahwah, NJ: Erlbaum.
- McGraw, B., & Jöreskog, K. G. (1971). Factorial invariance of ability measures in groups differing in intelligence and social economic status. *British Journal of Mathematical and Statistical Psychology*, 24, 154–168.
- Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, 5, 23–43.
- Miyake, A. (2001). Individual differences in working memory: Introduction to the special section. *Journal of Experimental Psychology*, 130, 163–168.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.
- Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of central executive. *British Journal of Psychology*, 91, 111–121.
- National Science Foundation. (2000). *National Survey of Science and Mathematics Education*. Reston, VA: Author.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 411–421.
- Oberauer, K., Suss, H., Wilhelm, O., Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, 31, 167–193.
- Ornstein, P., Naus, M. J., & Liberty, C. (1975). Rehearsal and organization processes in children’s memory. *Child Development*, 46, 818–830.
- Passolunghi, M. C., Cornoldi, C., & De Liberto, S. (2001). Working memory and intrusions of irrelevant information in a group of specific poor problem solvers. *Memory & Cognition*, 27, 779–790.
- Perfetti, C. (1985). *Reading ability*. New York: Oxford University Press.
- Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74, 525–556.
- Plewis, I. (1996). Statistical methods for understanding cognitive growth: A review, a synthesis and an application. *British Journal of Mathematical and Statistical Psychology*, 49, 25–42.
- Psychological Corporation. (1991). *Wechsler Intelligence Scale for Children—Third Edition*. San Antonio, TX: Harcourt Brace Jovanovich.
- Psychological Corporation. (1992). *Wechsler Individual Achievement Test*. San Antonio, TX: Harcourt Brace.
- Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology*, 91, 137–157.
- Raven, J. C. (1976). *Colored progressive matrices*. London: H. K. Lewis.
- Reise, S. P., Widaman, K., & Pugh, R. (1993). Confirmatory factor analysis and item response theory: Two approaches for explaining measurement invariance. *Psychological Bulletin*, 114, 552–566.
- Rende, B., Ramsberger, G., & Miyake, A. (2002). Commonalities and differences in working memory components underlying letter and category fluency tasks: A dual-task investigation. *Neuropsychology*, 16, 309–321.
- Riley, M., Greeno, J. G., & Heller, J. (1983). The development of children’s problem solving ability in arithmetic. In H. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York: Academic Press.
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, 93, 346–362.
- Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, 126, 211–227.
- SAS Institute. (2003). *SAS/STAT user’s guide* [Computer software manual]. Cary, NC: Author.
- Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Author.
- Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorder. *Cognition*, 24, 139–168.
- Siegler, R. S. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development*, 59, 833–851.
- Singer, J. D., (2002). Fitting individual growth models using SAS PROC MIXED. In Moskowitz & S. Hersberger (Eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications* (pp. 135–170). Mahwah, NJ: Erlbaum.
- Snijders, T. A., & Bosker, R. J. (2003). *Multilevel analysis*. Thousand Oaks, CA: Sage.
- Stoner, S. B. (1982). Age differences in crystallized and fluid intellectual abilities. *Journal of Psychology*, 110, 7–10.
- Swanson, H. L. (1992). Generality and modifiability of working memory among skilled and less skilled readers. *Journal of Educational Psychology*, 84, 473–488.
- Swanson, H. L. (1995). *S-Cognitive Processing Test*. Austin, TX: PRO-ED.
- Swanson, H. L. (1999). Reading comprehension and working memory in skilled readers: Is the phonological loop more important than the executive system? *Journal of Experimental Child Psychology*, 72, 1–31.
- Swanson, H. L. (2003). Age-related differences in learning disabled and skilled readers’ working memory. *Journal of Experimental Child Psychology*, 85, 1–31.
- Swanson, H. L. (2004). Working memory and phonological processing as predictors of children’s mathematical problem solving at different ages. *Memory & Cognition*, 32, 648–661.
- Swanson, H. L. (2006). Cross-sectional and incremental changes in working memory and mathematical problem solving. *Journal of Educational Psychology*, 98, 265–281.
- Swanson, H. L., & Ashbaker, M. (2000). Working memory, short-term memory, articulation speed, word recognition, and reading comprehension in learning disabled readers: Executive and/or articulatory system? *Intelligence*, 28, 1–30.
- Swanson, H. L., Ashbaker, M., & Lee, C. (1996). Learning disabled readers’ working memory as a function of processing demands. *Journal of Experimental Child Psychology*, 61, 242–275.
- Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not a risk for serious math difficulties. *Journal of Educational Psychology*, 96, 471–491.
- Swanson, H. L., & Berninger, V. (1995). The role of working memory and STM in skilled and less skilled readers’ word recognition and comprehension. *Intelligence*, 21, 83–108.
- Swanson, H. L., Cooney, J. B., & Brock, S. (1993). The influence of working memory and classification ability on children’s word problem solution. *Journal of Experimental Child Psychology*, 55, 374–395.
- Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review of Educational Research*, 76, 249–274.
- Swanson, H. L., Mink, J., & Bocian, K. M. (1999). Cognitive processing deficits in poor readers with symptoms of reading disabilities and ADHD: More alike than different? *Journal of Educational Psychology*, 91, 321–353.
- Swanson, H. L., & Sachse-Lee, C. (2001). Mathematic problem solving and working memory in children with learning disabilities: Both exec-

- utive and phonological processes are important. *Journal of Experimental Child Psychology*, 79, 294–321.
- Towse, J. (1998). On random generation and the central executive of working memory. *British Journal of Psychology*, 89, 77–101.
- Wagner, R., & Torgesen, J. (1999). *Test of Word Reading Efficiency*. Austin, TX: PRO-ED.
- Wagner, R., Torgesen, J., & Rashotte, C. (2000). *Comprehensive Test of Phonological Processes*. Austin, TX: PRO-ED.
- Whitney, P., Arnett, P., Driver, A., & Budd, D. (2001). Measuring central executive functioning: What's in a reading span? *Brain & Cognition*, 45, 1–14.
- Wilkinson, G. S. (1993). *The Wide Range Achievement Test—Third Edition*. Wilmington, DE: Wide Range.
- Wilson, R. S., Beckett, L. A., Barnes, L., Schneider, J., Bach, J., Evans, D., & Bennett, D. A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging*, 17, 179–193.
- Woodcock, R. W. (1998). *Woodcock Reading Mastery Test—Revised (Form G)*. Circle Pines, MN: American Guidance

Appendix A

Variable, Sample Size, Means, and Standard Deviations for Wave 1
(Grades 1, 2, and 3)

Variable	At risk for SMD			Not at risk for SMD		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Chronological age (in months)						
Age 1	77	80.09	4.55	54	80.04	5.38
Age 2	69	92.64	4.70	45	92.84	5.35
Age 3	64	105.42	7.69	44	104.09	4.85
Fluid intelligence (Raven)						
RAV-SS 1	77	102.29	15.99	54	111.07	12.78
RAV-SS 2	69	106.68	12.86	45	115.00	13.04
RAV-SS 3	64	105.42	14.20	44	118.89	10.86
RAV-R 1	77	18.62	5.30	54	22.06	5.39
RAV-R 2	69	22.96	5.40	45	26.53	5.67
RAV-R 3	64	25.69	5.75	44	30.57	3.91
Word problems (WISC-III)						
Arithm-SC 1	77	6.95	.80	54	12.94	2.22
Arithm-SC 2	69	8.65	1.53	45	11.02	3.43
Arithm-SC 3	64	8.92	1.86	44	11.18	2.37
Arithm-R 1	77	8.65	2.68	54	12.80	1.02
Arithm-R 2	69	11.83	2.20	45	13.36	2.13
Arithm-R 3	64	13.47	2.24	44	15.09	1.67
Digit naming speed (CTOPP)						
RDN-SC 1	77	6.99	2.72	54	8.51	1.61
RDN-SC 2	69	8.65	1.86	45	10.04	2.31
RDN-SC 3	64	8.92	3.09	44	11.38	2.47
RDN-R 1	77	59.31	14.75	54	43.72	8.95
RDN-R 2	69	44.54	8.80	45	36.47	8.76
RDN-R 3	64	37.83	7.52	44	31.93	7.62
Word problems: Semantic variation						
WPSV 1	77	2.06	1.66	54	3.04	1.61
WPSV 2	69	3.16	1.49	45	3.49	1.08
WPSV 3	64	3.78	1.33	44	4.09	1.93
Word problem solving processes (total of problem solving components)						
COMPON 1	77	6.77	2.95	54	8.67	2.56
COMPON 2	69	8.58	3.34	46	10.09	3.00
COMPON 3	64	10.39	3.03	43	12.60	2.03
Knowledge of problem solving components						
QUES 1	77	1.47	0.90	54	2.15	0.81
QUES 2	69	1.87	0.95	46	2.26	0.93
QUES 3	64	2.22	0.83	43	2.65	0.65
NUMB 1	77	1.53	1.06	54	2.15	0.98
NUMB 2	69	2.09	0.98	46	2.28	0.81
NUMB 3	64	2.56	0.73	43	2.81	0.45
Goal 1	77	1.19	0.84	54	1.15	0.92
Goal 2	69	0.97	0.86	46	1.39	0.95
Goal 3	64	1.81	0.97	43	2.40	0.69
OPER 1	77	1.44	0.75	54	1.61	0.81
OPER 2	69	1.90	0.99	46	2.15	0.82

(Appendixes continue)

Appendix A (continued)

Variable	At risk for SMD			Not at risk for SMD		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
OPER 3	64	1.58	1.12	43	2.09	1.02
ALGO 1	77	1.13	0.83	54	1.61	0.86
ALGO 2	69	1.75	0.95	46	2.00	0.79
ALGO 3	64	2.22	0.86	43	2.65	0.53
IRRE 1	77	1.86	0.81	54	2.00	0.73
IRRE 2	69	2.04	0.96	46	2.24	0.79
IRRE 3	64	2.45	0.85	43	2.86	0.41
Math calculation (WRAT, WIAT, CBM)						
AWRAT-SS 1	77	111.27	11.14	54	115.35	10.37
AWRAT-SS 2	69	101.17	11.36	45	106.29	9.93
AWRAT-SS 3	64	106.86	14.79	44	116.18	12.55
AWRAT-R 1	77	18.77	1.59	54	19.69	1.49
AWRAT-R 2	69	21.86	2.14	45	23.07	1.96
AWRAT-R 3	64	27.45	3.74	44	29.34	3.23
WIAT-SS 1	77	103.23	13.87	54	110.19	12.95
WIAT-SS 2	69	102.72	14.78	45	107.80	12.24
WIAT-SS 3	62	109.69	14.17	43	118.23	10.63
WIAT-R 1	77	8.44	2.30	54	10.07	1.87
WIAT-R 2	69	13.87	3.19	45	15.36	2.83
WIAT-R 3	62	19.63	3.59	43	22.05	2.54
CBM-R 1	77	26.81	12.03	54	28.39	11.25
CBM-R 2	69	26.48	11.59	45	27.82	10.00
CBM-R 3	64	25.95	16.12	43	30.79	12.73
Phonological knowledge (Elision CTOPP, Pseudowords TOWRE)						
EL-SC 1	77	8.88	3.23	54	12.13	3.29
EL-SC 2	69	9.17	3.21	45	11.27	3.16
EL-SC 3	64	9.03	3.09	44	11.09	2.94
EL-R 1	77	5.48	4.08	54	10.24	4.90
EL-R 2	69	8.87	4.79	45	12.24	4.90
EL-R 3	64	10.66	4.92	44	13.91	4.66
PW-SS 1	77	101.22	13.10	54	112.93	12.81
PW-SS 2	69	101.45	12.64	45	112.07	12.66
PW-SS 3	64	101.06	13.61	44	113.05	15.69
PW-R 1	77	9.78	8.04	54	18.69	10.18
PW-R 2	69	18.96	10.33	45	28.51	11.00
PW-R 3	64	25.23	12.06	44	35.18	12.33
Reading skill (WRAT-III, Real Words TOWRE, Comprehension WRMT)						
WREAD-SS 1	77	103.64	16.65	54	116.50	15.05
WREAD-SS 2	69	100.57	16.66	45	107.91	12.23
WREAD-SS 3	64	99.16	16.90	44	110.23	14.52
WREAD-R 1	77	21.23	4.60	54	25.61	4.24
WREAD-R 2	69	27.32	5.17	45	29.91	3.72
WREAD-R 3	64	30.08	5.87	44	34.07	5.25
SWE-SS 1	77	100.35	12.77	54	113.78	12.43
SWE-SS 2	69	101.58	17.48	45	115.31	12.62
SWE-SS 3	64	101.38	14.51	44	113.36	12.22
SWE-TC 1	77	21.94	13.24	54	38.33	15.14
SWE-TC 2	69	44.42	17.64	45	58.53	12.59
SWE-TC 3	64	53.17	17.09	44	66.30	11.61
WRMT-SS 1	77	101.77	13.21	54	112.22	10.45
WRMT-SS 2	69	101.45	14.20	45	110.36	8.77
WRMT-SS 3	64	98.08	10.79	44	107.91	8.49
WRMT-R 1	77	11.71	7.97	54	20.28	7.76
WRMT-R 2	69	21.94	10.10	45	29.24	6.19
WRMT-R 3	64	26.95	8.72	44	35.39	5.04
Letter naming speed (CTOPP)						
RLN-SC 1	77	9.09	1.60	54	10.85	1.57
RLN-SC 2	69	10.20	1.87	45	11.62	2.27
RLN-SC 3	64	10.77	3.88	44	11.59	3.07
RLN-R 1	77	72.09	23.90	54	49.33	9.51
RLN-R 2	69	51.68	11.80	45	41.38	9.06
RLN-R 3	64	43.95	10.08	44	37.23	9.00
Short-term memory (phonological loop)						
RWORD 1	77	6.61	2.88	54	8.37	3.28

Appendix A (continued)

Variable	At risk for SMD			Not at risk for SMD		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
RWORD 2	69	7.48	3.00	45	7.84	3.74
RWORD 3	64	7.70	3.12	44	9.00	3.38
PWORD 1	77	2.68	1.63	54	3.44	1.93
PWORD 2	69	2.45	1.41	45	2.76	1.48
PWORD 3	64	2.72	1.47	44	2.91	1.64
DIG-F 1	77	4.62	2.41	54	5.69	3.02
DIG-F 2	69	5.00	2.56	45	5.49	3.11
DIG-F 3	64	5.52	2.85	44	7.55	2.91
DIG-B 1	77	2.32	0.95	54	3.00	1.06
DIG-B 2	69	2.81	1.34	45	3.62	1.54
DIG-B 3	64	3.28	1.24	44	4.18	1.67
Working memory (executive)						
LISSPAN 1	77	2.51	3.08	54	4.04	5.17
LISSPAN 2	69	3.67	3.69	45	6.34	5.18
LISSPAN 3	61	6.88	5.01	43	9.56	5.69
SEMSET 1	77	3.62	2.88	54	4.63	3.56
SEMSET 2	69	5.35	5.15	45	8.49	6.39
SEMSET 3	64	6.06	4.61	44	7.66	6.29
AUDSET 1	77	3.79	3.25	54	4.48	3.60
AUDSET 2	69	5.75	4.62	45	4.58	3.96
AUDSET 3	64	7.44	5.38	44	11.91	7.38
UPDATE 1	77	2.61	3.53	54	5.37	4.89
UPDATE 2	64	3.81	3.67	44	6.11	4.45
UPDATE 3	64	4.03	3.75	45	6.09	4.32
Working memory: Visual-spatial						
MATRIX 1	77	4.94	3.39	54	5.31	3.26
MATRIX 2	69	5.97	3.01	45	7.24	3.20
MATRIX 3	64	7.52	3.15	44	8.93	4.27
MAP-DIR 1	77	3.77	2.63	54	3.26	1.67
MAP-DIR 2	69	2.33	2.49	45	3.76	4.73
MAP-DIR 3	64	3.19	3.80	44	5.64	5.72
Fluency						
CATF 1	77	9.66	4.19	54	12.59	4.46
CATF 2	69	11.19	4.93	45	14.42	4.50
CATF 3	64	13.00	4.12	44	15.18	4.76
LETF 1	77	5.47	3.02	54	6.59	3.67
LETF 2	69	6.72	3.46	45	7.76	3.28
LETF 3	64	7.47	3.47	44	9.16	3.65
Random generation						
RANDL 1	77	0.49	0.30	54	0.42	0.24
RANDL 2	67	0.51	0.28	45	0.44	0.23
RANDL 3	64	0.40	0.19	44	0.37	0.15
RANDN 1	77	0.23	0.18	54	0.24	0.17
RANDN 2	69	0.26	0.13	45	0.27	0.12
RANDN 3	64	0.27	0.52	44	0.23	0.09

Note. The numbers 1, 2, and 3 at the end of each abbreviation refer to Waves 1, 2, and 3, respectively. SC = scale score ($M = 10$, $SD = 2$); SS = standard score; R = Raw score; RAV = Raven Progressive Matrices Test; WISC-III = Wechsler Intelligence Scale for Children—Third Edition; Arithm = word problem for the WISC-III; CTOPP = Comprehensive Test of Phonological Processing; RDN = rapid digit naming; WPSV = word problems with semantic variations; COMPOS = total of knowledge of word problem components; QUES = question; NUMB = number; OPER = operations; ALGO = algorithm knowledge; IRRE = Irrelevant; WRAT = Wide Range Achievement Test; WIAT = Wechsler Individual Achievement Test; CBM = Curriculum; AWRA = arithmetic subtest of WRAT; TWORE = Test of Word Reading Efficiency; EL = Elision subtest (CTOPP); PW = pseudoword fluency (TOWRE); WRAT-III = Wide Range Achievement Test—Third Edition; WREAD = WRAT reading score; SWE = Real word fluency (TOWRE); WRMT = passage comprehension Woodcock Reading Mastery Test—Revised; RLN = rapid naming of letters; RWORD = real word recall; PWORD = pseudoword recall; DIG-F = digits forward task; DIG-B = digits backward task; LISSPAN = Listening Span task; SEMSET = Semantic Association Span task; AUDSET = auditory Digit/Sentence Span task; UPDATE = Updating Task; MATRIX = visual-spatial matrix task; MAP-DIR = mapping and direction task; CATF = categorical fluency; LETF = letter fluency; RANDL = random generation of letters; RANDN = random generation of numbers.

(Appendixes continue)

Appendix B

Variable, Sample Size, Means, and Standard Deviations Starting at Grade 2

Variable	At risk for SMD			Not at risk for SMD		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Chronological age (in months)						
Age 1	34	94.06	5.25	58	92.79	3.88
Age 2	32	106.13	4.95	53	105.43	4.16
Age 3	30	117.80	4.82	53	117.15	5.20
Fluid intelligence (Raven)						
RAV-SS 1	34	104.18	11.33	58	111.98	15.91
RAV-SS 2	32	105.91	12.78	53	112.74	12.67
RAV-SS 3	30	107.27	10.93	53	111.91	11.23
RAV-R 1	34	23.41	4.86	58	25.86	5.83
RAV-R 2	32	26.50	5.18	53	28.98	4.63
RAV-R 3	30	29.40	3.87	53	30.72	3.81
Word problems (WISC-III)						
Arithm-SC 1	33	8.00	1.80	58	11.76	2.18
Arithm-SC 2	32	9.09	3.11	53	11.74	2.59
Arithm-SC 3	30	8.97	2.77	53	11.02	2.68
Arithm-R 1	33	11.79	1.73	58	14.19	1.19
Arithm-R 2	32	13.81	2.09	53	15.74	1.84
Arithm-R 3	30	15.10	1.99	53	16.66	2.08
Digit naming speed (CTOPP)						
RDN-SC 1	33	7.78	1.01	58	9.35	1.53
RDN-SC 2	32	8.84	1.64	53	10.82	2.00
RDN-SC 3	30	10.06	2.04	53	11.86	2.16
RDN-R 1	33	48.76	10.62	58	37.91	6.33
RDN-R 2	32	41.91	10.13	53	33.17	6.47
RDN-R 3	30	36.53	7.66	53	30.40	6.22
Word problems: Semantic variation						
WPSV 1	33	3.21	1.24	58	3.78	1.23
WPSV 2	32	3.75	1.02	53	3.96	1.19
WPSV 3	30	4.17	1.26	53	4.51	1.61
Word problem solving processes (total of problem solving components)						
COMPON 1	34	8.38	3.08	58	10.47	2.54
COMPON 2	32	11.53	2.14	53	12.79	2.01
COMPON 3	30	16.40	2.14	53	17.60	2.61
Knowledge of problem solving components						
QUES 1	34	1.62	0.85	58	1.98	0.87
QUES 2	32	2.28	1.02	53	2.62	0.60
QUES 3	30	3.23	1.04	53	3.58	0.66
NUMB 1	34	2.06	0.89	58	2.43	0.68
NUMB 2	32	2.50	0.62	53	2.79	0.45
NUMB 3	30	3.50	0.63	53	3.75	0.52
Goal 1	34	0.85	0.86	58	1.45	0.98
Goal 2	32	2.09	0.82	53	2.28	0.91
Goal 3	30	3.07	0.83	53	3.25	0.98
OPER 1	34	1.91	1.03	58	2.41	0.77
OPER 2	32	2.19	0.54	53	2.43	0.69
OPER 3	30	3.13	0.51	53	3.36	0.88
ALGO 1	34	1.94	0.92	58	2.19	0.71
ALGO 2	32	2.47	0.62	53	2.66	0.59
ALGO 3	30	3.47	0.63	53	3.66	0.62
IRRE 1	34	2.03	0.90	58	2.40	0.70
IRRE 2	32	2.59	0.71	53	2.79	0.53
IRRE 3	30	2.83	0.46	53	2.70	0.61
Math calculation (WRAT, WIAT, CBM)						
AWRAT-SS 1	34	99.59	9.97	58	106.12	9.62
AWRAT-SS 2	32	107.00	16.13	53	118.55	13.26
AWRAT-SS 3	30	106.93	14.29	53	116.49	10.61
AWRAT-R 1	34	21.94	1.72	58	23.34	1.87
AWRAT-R 2	32	27.94	4.14	53	31.00	3.24
AWRAT-R 3	30	31.37	4.06	53	34.13	2.76
WIAT-SS 1	34	102.06	12.13	58	108.86	10.55
WIAT-SS 2	32	111.13	15.50	53	119.96	13.45
WIAT-SS 3	30	108.23	15.49	53	117.36	11.32

Appendix B (continued)

Variable	At risk for SMD			Not at risk for SMD		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
WIAT-R 1	34	14.29	2.74	58	15.72	2.14
WIAT-R 2	32	20.56	3.75	53	22.94	3.17
WIAT-R 3	30	23.97	4.79	53	26.77	2.94
CBM-R 1	34	27.35	18.51	58	27.48	11.26
CBM-R 2	32	37.75	17.52	53	33.36	11.60
CBM-R 3	30	70.10	20.45	53	75.66	26.23
Phonological knowledge (Elision CTOPP, Pseudowords TOWRE)						
EL-SC 1	33	9.45	3.34	58	11.17	3.15
EL-SC 2	32	8.94	2.38	53	10.64	3.16
EL-SC 3	30	8.90	2.83	53	11.30	2.95
EL-R 1	33	9.82	5.29	58	12.28	4.89
EL-R 2	32	10.84	4.06	53	13.28	4.92
EL-R 3	30	12.10	4.57	53	15.68	4.24
PW-SS 1	33	101.45	11.73	58	111.26	12.47
PW-SS 2	32	100.84	12.49	53	111.13	13.59
PW-SS 3	30	97.23	13.04	53	106.68	12.25
PW-R 1	33	18.70	9.82	58	27.48	10.86
PW-R 2	32	24.59	10.73	53	33.60	10.62
PW-R 3	30	26.77	11.88	53	35.45	10.26
Reading skill (WRAT-III, Real Words TOWRE, Comprehension WRMT)						
WREAD-SS 1	33	98.70	12.91	58	108.41	12.83
WREAD-SS 2	32	97.31	12.85	53	106.83	13.23
WREAD-SS 3	30	99.83	15.03	53	110.00	12.85
WREAD-R 1	33	27.36	4.22	58	30.76	4.27
WREAD-R 2	32	29.88	4.19	53	33.25	4.51
WREAD-R 3	30	33.13	5.14	53	36.85	4.40
SWE-SS 1	33	103.18	13.34	58	115.47	12.16
SWE-SS 2	32	102.97	11.33	53	113.13	10.12
SWE-SS 3	30	99.57	11.74	53	108.96	10.36
SWE-TC 1	33	44.42	13.21	58	58.12	10.40
SWE-TC 2	32	55.94	12.18	53	66.25	8.95
SWE-TC 3	30	62.17	12.65	53	71.32	9.39
WRMT-SS 1	33	103.03	10.24	58	109.07	10.70
WRMT-SS 2	32	98.84	8.69	53	106.87	9.66
WRMT-SS 3	30	97.60	7.41	53	104.57	10.55
WRMT-R 1	33	25.00	6.60	58	29.84	5.96
WRMT-R 2	32	29.34	5.94	53	34.45	5.66
WRMT-R 3	30	33.47	4.85	53	37.49	6.32
Letter naming speed (CTOPP)						
RLN-SC 1	33	9.33	1.80	58	10.93	1.99
RLN-SC 2	32	9.47	2.00	53	10.79	2.11
RLN-SC 3	30	9.70	2.37	53	10.75	2.03
RLN-R 1	33	52.52	12.69	58	42.88	7.97
RLN-R 2	32	45.78	10.62	53	39.04	8.60
RLN-R 3	30	39.80	8.81	53	35.06	6.52
Short-term memory (phonological loop)						
RWORD 1	33	8.09	3.51	58	9.53	3.15
RWORD 2	32	7.59	3.66	53	9.60	3.41
RWORD 3	30	8.50	2.96	53	10.66	3.51
PWORD 1	33	3.70	1.85	58	3.90	2.24
PWORD 2	32	2.59	1.78	53	3.21	1.79
PWORD 3	30	3.23	1.91	53	2.98	1.74
DIG-F 1	33	6.03	2.82	58	6.84	2.73
DIG-F 2	32	5.69	3.01	53	7.09	3.36
DIG-F 3	30	6.20	3.27	53	8.34	2.59
DIG-B 1	33	2.82	0.98	58	3.43	1.30
DIG-B 2	32	3.34	1.23	53	4.13	1.87
DIG-B 3	30	3.50	1.46	53	4.66	2.39

(Appendixes continue)

Appendix B (continued)

Variable	At risk for SMD			Not at risk for SMD		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Working memory (executive)						
LISSPAN 1	33	4.88	4.04	58	6.59	4.95
LISSPAN 2	31	5.63	4.89	52	8.65	4.62
LISSPAN 3	30	8.25	4.04	52	10.36	4.64
SEMSET 1	33	5.27	4.22	58	4.60	3.55
SEMSET 2	32	6.09	5.11	51	7.88	6.28
SEMSET 3	30	7.57	5.83	53	9.08	7.98
AUDSET 1	33	4.64	5.02	58	5.45	4.74
AUDSET 2	32	6.75	4.20	53	11.11	7.10
AUDSET 3	30	10.87	6.98	53	13.81	9.15
UPDATE 1	33	3.94	3.66	58	7.00	5.06
UPDATE 2	30	4.47	3.78	53	6.57	4.58
UPDATE 3	30	4.60	3.77	51	6.98	4.92
Working memory: Visual-spatial						
MATRIX 1	34	5.97	4.12	58	6.83	3.74
MATRIX 2	32	6.84	3.33	53	7.75	4.55
MATRIX 3	30	8.47	4.20	53	11.13	5.95
MAP-DIR 1	33	3.42	4.36	58	3.76	2.38
MAP-DIR 2	32	4.25	5.10	53	3.94	4.80
MAP-DIR 3	30	7.33	6.91	53	6.74	6.29
Fluency						
CATF 1	33	12.76	3.98	58	13.86	3.96
CATF 2	32	13.03	3.94	53	15.64	4.68
CATF 3	30	14.50	3.91	53	16.42	4.47
LETF 1	33	7.09	2.79	58	7.07	3.09
LETF 2	32	7.56	2.83	53	8.21	3.26
LETF 3	30	8.43	2.85	53	9.21	4.78
Random generation						
RANDL 1	34	0.48	0.27	58	0.45	0.22
RANDL 2	32	0.50	0.23	53	0.48	0.20
RANDL 3	30	0.39	0.16	53	0.36	0.12
RANDN 1	34	0.15	0.10	58	0.23	0.15
RANDN 2	32	0.24	0.22	53	0.24	0.11
RANDN 3	30	0.23	0.11	53	0.26	0.18

Note. The numbers 1, 2, and 3 at the end of each abbreviation refer to Waves 1, 2, and 3, respectively. SC = scale score ($M = 10$, $SD = 2$); SS = standard score; R = Raw score; RAV = Raven Progressive Matrices Test; WISC-III = Wechsler Intelligence Scale for Children—Third Edition; Arithm = word problem for the WISC-III; CTOPP = Comprehensive Test of Phonological Processing; RDN = rapid digit naming; WPSV = word problems with semantic variations; COMPON = total of knowledge of word problem components; QUES = question; NUMB = number; OPER = operations; ALGO = algorithm knowledge; IRRE = Irrelevant; WRAT = Wide Range Achievement Test; WIAT = Wechsler Individual Achievement Test; CBM = Curriculum; AWRAT = arithmetic subtest of WRAT; TWORE = Test of Word Reading Efficiency; EL = Elision subtest (CTOPP); PW = pseudoword fluency (TOWRE); WRAT-III = Wide Range Achievement Test—Third Edition; WREAD = WRAT reading score; SWE = Real word fluency (TOWRE); WRMT = passage comprehension Woodcock Reading Mastery Test—Revised; RLN = rapid naming of letters; RWORD = real word recall; PWORD = pseudoword recall; DIG-F = digits forward task; DIG-B = digits backward task; LISSPAN = Listening Span task; SEMSET = Semantic Association Span task; AUDSET = auditory Digit/Sentence Span task; UPDATE = Updating Task; MATRIX = visual-spatial matrix task; MAP-DIR = mapping and direction task; CATF = categorical fluency; LETF = letter fluency; RANDL = random generation of letters; RANDN = random generation of numbers.

Appendix C

Variable, Sample Size, Means, and Standard Deviations Starting at Grade 3

Variable	At risk for SMD			Not at risk for SMD		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Chronological age (in months)						
Age 1	24	107.88	6.35	107	104.37	4.91
Age 2	22	120.82	6.61	98	116.60	4.09
Age 3	22	132.73	6.27	89	128.18	6.71
Fluid intelligence (Raven)						
RAV-SS 1	24	96.71	9.70	107	108.47	13.42
RAV-SS 2	22	99.95	10.96	98	110.36	11.57
RAV-SS 3	22	102.27	10.17	88	109.07	13.81
RAV-R 1	24	23.79	4.55	107	27.67	5.00
RAV-R 2	22	27.00	4.35	98	30.09	4.04
RAV-R 3	22	29.95	3.54	88	31.25	3.95
Word problems (WISC-III)						
Arithm-SC 1	22	6.09	1.95	107	11.70	2.08
Arithm-SC 2	22	8.73	3.34	98	10.77	3.07
Arithm-SC 3	21	7.05	1.94	89	10.80	3.49
Arithm-R 1	22	11.91	2.51	107	15.74	1.57
Arithm-R 2	22	14.86	1.52	98	16.46	2.10
Arithm-R 3	21	14.90	1.61	89	17.79	3.26
Digit naming speed (CTOPP)						
RDN-SC 1	22	9.25	1.44	106	11.03	2.24
RDN-SC 2	22	10.77	2.10	98	12.04	2.49
RDN-SC 3	21	11.83	2.55	89	12.68	2.21
RDN-R 1	22	38.73	7.08	107	32.66	6.84
RDN-R 2	22	33.55	6.51	98	29.97	6.92
RDN-R 3	21	30.48	6.96	89	28.09	5.99
Word problems: Semantic variation						
WPSV 1	22	3.23	1.57	107	3.68	1.29
WPSV 2	22	3.73	1.49	98	4.18	1.12
WPSV 3	21	3.71	0.72	89	4.17	1.36
Word problem solving processes (total of problem solving components)						
COMPON 1	24	9.54	2.98	107	11.91	2.58
COMPON 2	22	15.05	2.61	98	17.29	2.43
COMPON 3	22	20.05	2.61	89	22.35	2.50
Knowledge of problem solving components						
QUES 1	24	1.79	1.06	107	2.45	0.78
QUES 2	22	2.82	0.80	98	3.47	0.71
QUES 3	22	3.82	0.80	89	4.48	0.71
NUMB 1	24	2.58	0.58	107	2.61	0.59
NUMB 2	22	3.45	0.67	98	3.56	0.59
NUMB 3	22	4.45	0.67	89	4.55	0.60
Goal 1	24	1.25	0.99	107	2.19	0.92
Goal 2	22	2.64	0.79	98	3.16	0.77
Goal 3	22	3.64	0.79	89	4.18	0.79
OPER 1	24	1.92	0.97	107	2.30	0.84
OPER 2	22	3.05	0.84	98	3.50	0.71
OPER 3	22	4.05	0.84	89	4.53	0.69
ALGO 1	24	2.00	0.88	107	2.36	0.73
ALGO 2	22	3.09	0.87	98	3.59	0.66
ALGO 3	22	4.09	0.87	89	4.61	0.65
IRRE 1	24	2.50	0.66	107	2.63	0.64
IRRE 2	22	2.73	0.55	98	2.86	0.35
IRRE 3	22	2.36	0.90	88	2.73	0.56
Math calculation (WRAT, WIAT, CBM)						
AWRAT-SS 1	24	106.79	14.83	107	117.61	12.09
AWRAT-SS 2	22	105.77	11.99	98	117.88	9.45
AWRAT-SS 3	22	99.36	13.59	88	113.72	11.10

(Appendixes continue)

Appendix C (continued)

Variable	At risk for SMD			Not at risk for SMD		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
AWRAT-R 1	24	28.25	3.64	107	30.13	2.89
AWRAT-R 2	22	31.68	3.14	98	34.27	2.43
AWRAT-R 3	22	32.77	3.66	88	36.32	3.54
WIAT-SS 1	24	108.17	13.74	107	116.63	9.94
WIAT-SS 2	22	108.32	12.63	98	119.68	9.88
WIAT-SS 3	22	98.41	14.70	88	111.34	10.04
WIAT-R 1	24	20.17	2.90	107	21.66	2.11
WIAT-R 2	22	25.05	3.30	98	27.35	2.81
WIAT-R 3	22	25.86	3.89	88	28.88	3.10
CBM-R 1	24	31.75	11.48	107	35.18	12.87
CBM-R 2	22	79.05	53.72	98	82.18	28.60
CBM-R 3	22	88.86	63.09	88	90.34	61.69
Phonological knowledge (Elision CTOPP, Pseudowords TOWRE)						
EL-SC 1	22	7.86	4.14	107	10.85	3.52
EL-SC 2	22	7.68	3.17	98	10.79	2.90
EL-SC 3	21	8.14	3.32	89	10.64	2.66
EL-R 1	22	9.59	5.74	107	13.53	5.41
EL-R 2	22	10.68	4.63	98	15.00	4.34
EL-R 3	21	13.14	4.92	89	16.25	3.84
PW-SS 1	22	96.18	11.16	107	110.74	13.42
PW-SS 2	22	99.64	13.83	98	108.74	14.22
PW-SS 3	21	93.38	13.23	89	104.72	13.41
PW-R 1	22	22.27	10.13	107	33.63	10.59
PW-R 2	22	29.27	12.57	98	36.90	11.43
PW-R 3	21	30.76	12.16	89	40.11	10.48
Reading skill (WRAT-III, Real Words TOWRE, Comprehension WRMT)						
WREAD-SS 1	22	94.32	16.70	107	107.18	11.51
WREAD-SS 2	22	94.68	16.46	98	107.39	12.16
WREAD-SS 3	21	91.81	16.78	89	108.94	12.71
WREAD-R 1	22	29.59	5.34	107	33.41	4.10
WREAD-R 2	22	32.09	4.97	98	35.74	4.37
WREAD-R 3	21	33.43	5.64	89	38.89	4.67
SWE-SS 1	22	99.23	14.72	107	112.29	10.38
SWE-SS 2	22	99.55	11.68	98	109.65	10.84
SWE-SS 3	21	94.24	11.57	89	105.53	9.92
SWE-TC 1	22	53.64	15.77	107	65.79	8.99
SWE-TC 2	22	61.68	13.05	98	71.68	10.09
SWE-TC 3	21	64.86	12.63	89	76.20	9.19
WRMT-SS 1	22	95.32	12.42	107	105.87	8.90
WRMT-SS 2	22	90.95	9.45	98	103.93	9.40
WRMT-SS 3	21	90.29	9.50	89	102.18	10.51
WRMT-R 1	22	28.14	8.20	107	34.27	4.60
WRMT-R 2	22	29.32	5.49	98	36.92	5.29
WRMT-R 3	21	33.67	5.76	89	40.09	6.44
Letter naming speed (CTOPP)						
RLN-SC 1	22	9.64	1.92	106	11.10	2.08
RLN-SC 2	22	9.77	2.51	98	11.10	2.65
RLN-SC 3	21	9.90	3.10	89	11.37	4.11
RLN-R 1	22	42.77	9.00	106	37.93	7.21
RLN-R 2	22	38.09	9.40	98	35.03	7.34
RLN-R 3	21	35.05	10.62	89	32.54	7.24
Short-term memory (phonological loop)						
RWORD 1	22	7.86	2.95	107	9.16	3.08
RWORD 2	22	7.45	3.13	98	10.03	3.81
RWORD 3	21	7.71	2.43	88	10.88	3.63
PWORD 1	22	2.95	1.62	107	3.82	1.96
PWORD 2	22	2.59	1.37	98	3.47	1.93
PWORD 3	21	2.95	1.69	89	3.51	1.61
DIG-F 1	22	4.68	2.12	107	7.22	3.09
DIG-F 2	22	5.86	2.71	98	6.94	3.00
DIG-F 3	21	7.29	2.53	89	8.20	3.00
DIG-B 1	22	2.82	1.05	107	3.61	1.61
DIG-B 2	22	3.50	1.74	98	4.07	1.83
DIG-B 3	21	3.62	1.28	89	4.62	1.78

Appendix C (continued)

Variable	At risk for SMD			Not at risk for SMD		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Working memory (executive)						
LISSPAN 1	22	4.02	4.25	107	6.77	4.06
LISSPAN 2	22	4.73	5.80	94	8.94	4.86
LISSPAN 3	21	7.60	5.40	86	11.31	5.08
SEMSET 1	22	4.64	3.55	106	6.54	5.34
SEMSET 2	22	7.23	6.02	97	9.68	7.11
SEMSET 3	21	6.86	4.56	88	11.86	8.81
AUDSET 1	22	4.68	3.85	107	7.87	6.13
AUDSET 2	22	9.27	5.36	98	12.88	7.94
AUDSET 3	21	11.71	6.68	89	14.61	8.92
UPDATE 1	22	2.95	3.24	107	8.24	5.19
UPDATE 2	21	5.76	4.36	89	7.78	4.42
UPDATE 3	20	5.50	4.30	91	7.44	4.40
Working memory: Visual-spatial						
MATRIX 1	24	7.04	3.61	107	7.93	4.65
MATRIX 2	22	8.50	4.01	98	9.20	5.63
MATRIX 3	22	9.45	5.35	88	10.97	6.28
MAP-DIR 1	22	3.91	3.08	107	5.05	4.63
MAP-DIR 2	22	4.09	4.77	97	5.66	6.54
MAP-DIR 3	21	4.81	5.36	89	7.34	6.89
Fluency						
CATF 1	22	13.23	3.98	107	15.35	4.27
CATF 2	22	14.09	4.77	98	16.88	5.24
CATF 3	21	14.43	4.08	89	16.92	5.17
LETF 1	22	7.05	2.89	107	8.39	3.39
LETF 2	22	9.36	5.09	98	8.88	3.47
LETF 3	21	8.76	4.00	89	10.25	3.48
Random generation						
RANDL 1	24	0.54	0.25	107	0.40	0.18
RANDL 2	22	0.47	0.21	98	0.38	0.19
RANDL 3	22	0.37	0.09	88	0.33	0.13
RANDN 1	24	0.17	0.13	107	0.20	0.11
RANDN 2	22	0.19	0.08	98	0.22	0.09
RANDN 3	22	0.19	0.07	88	0.20	0.11

Note. The numbers 1, 2, and 3 at the end of each abbreviation refer to Waves 1, 2, and 3, respectively. SC = scale score ($M = 10$, $SD = 2$); SS = standard score; R = Raw score; RAV = Raven Progressive Matrices Test; WISC-III = Wechsler Intelligence Scale for Children—Third Edition; Arithm = word problem for the WISC-III; CTOPP = Comprehensive Test of Phonological Processing; RDN = rapid digit naming; WPSV = word problems with semantic variations; COMON = total of knowledge of word problem components; QUES = question; NUMB = number; OPER = operations; ALGO = algorithm knowledge; IRRE = Irrelevant; WRAT = Wide Range Achievement Test; WIAT = Wechsler Individual Achievement Test; CBM = Curriculum; AWRAT = arithmetic subtest of WRAT; TWORE = Test of Word Reading Efficiency; EL = Elision subtest (CTOPP); PW = pseudoword fluency (TOWRE); WRAT-III = Wide Range Achievement Test—Third Edition; WREAD = WRAT reading score; SWE = Real word fluency (TOWRE); WRMT = passage comprehension Woodcock Reading Mastery Test—Revised; RLN = rapid naming of letters; RWORD = real word recall; PWORD = pseudoword recall; DIG-F = digits forward task; DIG-B = digits backward task; LISSPAN = Listening Span task; SEMSET = Semantic Association Span task; AUDSET = auditory Digit/Sentence Span task; UPDATE = Updating Task; MATRIX = visual-spatial matrix task; MAP-DIR = mapping and direction task; CATF = categorical fluency; LETF = letter fluency; RANDL = random generation of letters; RANDN = random generation of numbers.

Received August 24, 2006
Revision received August 21, 2007
Accepted August 28, 2007 ■